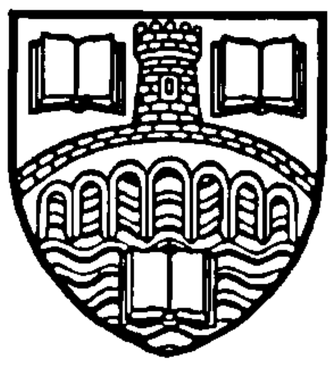
The Visual Representation of Texture

S.C. Dakin

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy

May 20, 1994

Supervisor: R.J. Watt

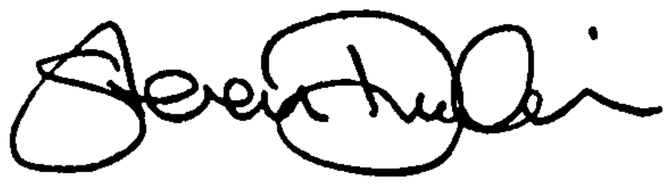


Centre for Cognitive and Computational Neuroscience
University of Stirling, Scotland FK9 4LA

Declaration

I declare that this thesis has been composed by myself, that the research reported therein has been conducted by myself unless otherwise stated, and that this work has not been included in another thesis

Stirling, May 20, 1994.



Steven Dakin

Acknowledgements

I am grateful to Roger Watt for offering me the opportunity to work in what has proved to be a unique environment, and for giving me the freedom, resources, and support that I needed to do this PhD.

Ian Paterson helped to develop many of the ideas underlying this thesis, and co-wrote much of the software that ran experiments and simulations. I thank him for his consistent enthusiasm for my work, and his generosity with his own ideas and time.

As well as frequently discussing my work with me, Dan Sturdy also proof-read the first draft of this thesis. His depressingly numerous comments helped to translate the draft into a closer approximation to English. Thanks to Dan. Thanks also to Roland Baddeley for many hours spent discussing texture and vision. His comments were consistently interesting and useful.

I also thank Will Goodall, Steve Emmott and Trish Carlin for sharing PhD-type traumas. Both Steve and Trish have discussed many aspects of my thesis with me, and helped me much more than they may realise and/or admit. Will Goodall, on the other hand, rarely talked about my thesis, but did give me the detailed technical assistance with computer equipment that no manuals could provide. Without the company and assistance of these three exceptional individuals I probably would have finished about six months ago.

The Centre for Computational and Cognitive Neuroscience is a unique working environment not because of its panoramic views or big computers, but because of: Ben Craven, Lawrence Gerstley, Jim Kay, Paul Miller, Todd Newton, Paul Toombs, and Tony Zawadski. All of them have made time to discuss my work with me and I thank them all for their patience and assistance. I am also grateful to Peter Hancock for technical support, and to Ric Lister for all the help with LaTeX.

Finally, Rachel O'Sullivan (Figure 1.3b) has tolerated my lousy moods, anti-social working habits, and unsavoury late-night-kitchen-related behaviour. She has listened to me drone on about this thesis at tedious length, and even served as a subject in some of the experiments. I thank her for making this time at Stirling the best.

This work was funded by the S.E.R.C.

Abstract

This research is concerned with texture: a source of visual information, that has motivated a huge amount of psychophysical and computational research. This thesis questions how useful the accepted view of texture perception is. From a theoretical point of view, work to date has largely avoided two critical aspects of a computational theory of texture perception. Firstly, *what is texture*? Secondly, what is an appropriate *representation* for texture? This thesis argues that a task dependent definition of texture is necessary, and proposes a multi-local, statistical scheme for representing texture orientation.

Human performance on a series of psychophysical orientation discrimination tasks are compared to specific predictions from the scheme.

The first set of experiments investigate observers' ability to directly derive statistical estimates from texture. An analogy is reported between the way texture statistics are derived, and the visual processing of spatio-luminance features.

The second set of experiments are concerned with the way texture elements are extracted from images (an example of the generic grouping problem in vision). The use of highly constrained experimental tasks, typically texture orientation discriminations, allows for the formulation of simple statistical criteria for setting critical parameters of the model (such as the spatial scale of analysis). It is shown that schemes based on isotropic filtering and symbolic matching do not suffice for performing this grouping, but that the scheme proposed, base on oriented mechanisms, does.

Taken together these results suggest a view of visual texture processing, not as a disparate collection of processes, but as a general strategy for deriving statistical representations of images common to a range of visual tasks.

Contents

1 | Introduction

1.1 Overview

This chapter introduces the problems considered in this thesis, in the context of a working definition of visual texture. Various definitions of texture are described, and it is proposed that the evident lack of consensus between these definitions is due to the fact that texture can only be defined *operationally*, i.e. in the context of a particular visual task. Section 1.3 expands this idea into a working definition based on statistical image description, describes necessary assumptions, and indicates where these assumptions are tested within the thesis. The literature concerning texture segmentation, shape from texture and texture flow, is then briefly reviewed. Finally, an overview of the structure of the thesis is given in Section 1.5.

1.2 What is texture?

A logical starting point for examining texture perception is to try and clearly define the class of stimulus being examined. It is interesting to note that, in the many papers written about the perception of visual texture, clear definitions of texture are few and far between. One is often told that "shape from texture" or "texture segmentation" is the subject of enquiry, but there does not seem to be a consensus of opinion as to what visual texture is. Three broad categories of definition have been offered, based on: surface quality, spatial scale, and statistics.

Surface quality:

> "Arrangement of threads, etc. in textile fabric, characteristic feel due to this; arrangement of small constituent parts, perceived structure, (*of* skin, rock, soil, organic tissue, literary work, etc.;" (Concise Oxford Dictionary)
>
> "...a term for the quality of a surface. The feature that dominates a texture scene is the repetitive or quasi-repetitive pattern." (Fu, 1974)
>
> "...the visual character of the surface of things." (Watt, 1988)
>
> "...the character of an object resulting from the arrangement or qualities of its particles or constituent parts." (Bergen, 1991)

Spatial scale:

> "...detailed structure in an image that is too fine to be resolved, yet coarse enough to produce noticeable fluctuation in the grey levels of neighbouring cells." (Horn, 1986)
>
> "..any visible features of a surface that are homogeneous in size and spacing across the extent of the surface... A strand of thread, a pebble, or a tree each has textured surfaces of its own, however, so that the visible environment may be thought of as containing nested hierarchies of textured surfaces." (Sedgwick, 1983; summarising Gibson, 1950a)
>
> "...the structure of a surface, as distinguished from the structure of the substance underlying the surface. We are talking about the relatively fine structure of the environment at the size-level around centimeters or millimeters."

(Gibson, 1979, p25)

"...an attribute of a field having no components that appear enumerable," (Richards and Pollit, 1988)

Statistics:

"(If you can) "see" the spaces between the individual droplets (of a cloud). Then you describe the microstructure as a *texture*. Thus you prepare only a statistical description." (Koenderink, 1990)

"Since texture refers to the *molding* of a surface, which may be irregular or regular, one might believe that a detailed description of the configuration of the surface would be equivalent to a description of its texture but this would not be correct. Although texture can be derived from the detailed microscopic knowledge properties of a surface... texture is not itself a microscopic property. Rather, it is... a *statistical property* of certain characteristic features of the surface." (Resnikoff, 1989)

There exists a fourth type of definition which amounts to a description of the experimental stimuli employed:

"Visual textures are defined as aggregates of many small elements. The elements can either be dots of certain colors...or simple patterns." (Julesz, 1991)

These classes of definition seem to address different *aspects* of visual texture processing. Given that that the visual processing of texture is a complex information processing task, a useful approach to examining such a system is that of Marr (1982). He proposed that there are three levels at which information processing systems can be understood. The first of those is the level of computational theory. At this level the goal of a computation and the reasons for its importance are stated. The second level is that of the algorithm. This states how the computational theory can be implemented, along with appropriate representations for the input and output. The final stage is that of the hardware implementation, which describes the physical realisation of the proposed algorithm. Levels mutually constrain each other: the goals outlined in the computational theory constrain the algorithm, which is in turn constrained by what hardware is available to perform the task.

Consider the first set of definitions, based on surface "character". By "character" the authors do not refer to traits and qualities that would uniquely identify a particular *instance* of a texture. Instead they refer to traits that place the texture into some broader category, e.g. wood, stone, etc. What are these traits? Typically they are attributes of the object that reflect how it was formed, and which consequently constrain how we can behave towards it. For example, the oriented structure of woodgrain reflects the multiple forces occurring in the growth of a tree, which in turn constrains how we might cut the wood to avoid it splintering. This is an important point: we perceive texture for a purpose, to better define the shape of an object, to determine if an object will afford sufficient grip to allow us to pick it up, to determine if the face of a mountain can be climbed in a day, etc. Surface traits capture some aspect of the *active* nature of texture. Indeed, this level of definition is probably closest to Marr's computational theory, being a statement of the intended *goal* of a process.

The second set of definitions refer to the spatial scale of objects and the texture that covers them. These definitions presuppose that there is a correct scale to consider objects, and a correct (finer) scale to consider texture. They also constrain the type of surface qualities that texture is concerned with. Because the elements of texture are at a finer scale than the surface they cover, emergent qualities must result from some form of aggregation of those elements over space. This follows on to the final set of definitions, which describe texture not as a surface property, but as a statistical property. A description such as this is closer to a description of the representation of texture (i.e. Marr's algorithmic level). The definitions relating to spatial scale and statistics are both *algorithmic* levels of explanation. They describe aspects of the strategies (i.e. operation at different spatial scales) and representations (statistics) that texture vision uses to achieve the goals described in the computational theory.

To summarise, definitions of texture vary because texture can only be defined operationally, and because authors have attempted to describe this operation at different levels of information processing. For that reason the following argument will primarily refer to texture processing, i.e. tasks that operate on texture.

1.3 An alternative view of texture

In this section texture processing is considered as an operation common to a number of visual tasks, whose essence is the calculation of statistical image descriptions. The section

is organised into a set of descriptions of the key assumptions underlying this view, each followed by a short summary of any pertinent evidence.

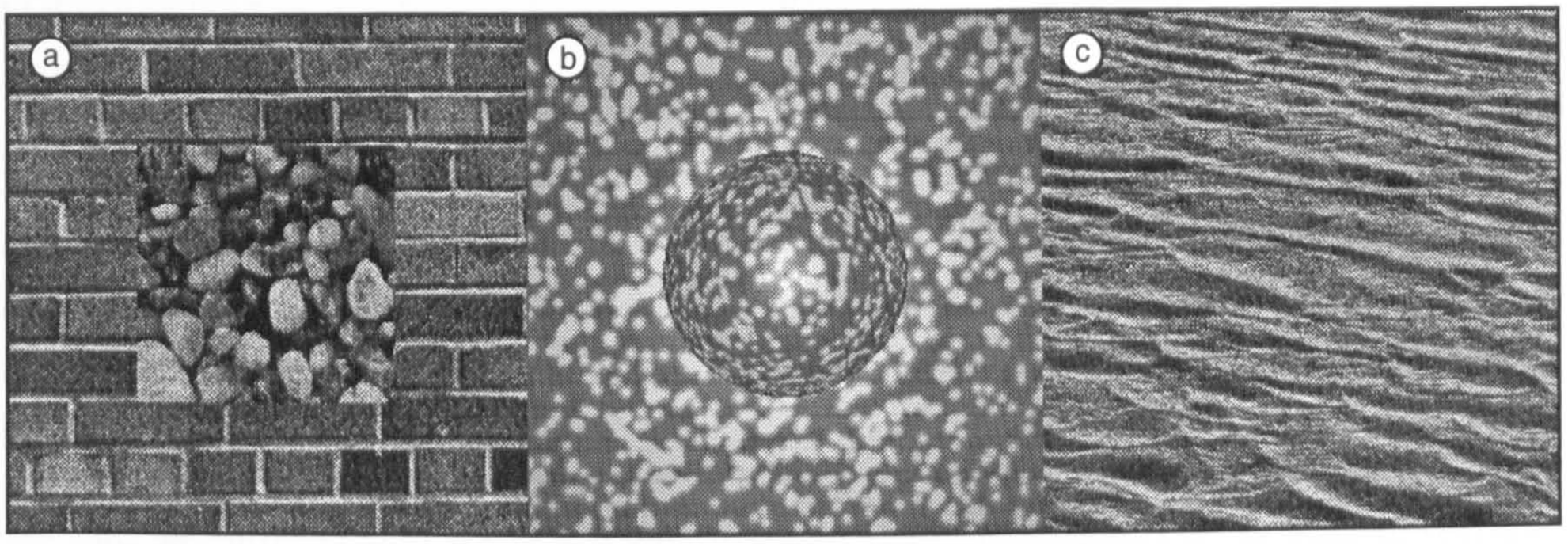


Figure 1.1. Examples of (a) texture segmentation, (b) shape from texture, (c) texture flow.

Assumption 1 *Texture processing is a component of a number of visual tasks.*

Texture information is used for a variety of visual tasks: discontinuities signal surface discontinuities (e.g. Figure 1.1a; Julesz, 1981), variations in texture geometry signal local shape (e.g. Figure 1.1b; Gibson, 1950b), texture information reflects the formative processes a surface has undergone (Figure 1.1c; Witkin and Tennenbaum, 1983), texture may be coded as component of object recognition, etc. Such visual operations which makes use of texture information will be referred to as "texture tasks". Each task has its own goals and constraints, but shares a common sets of operations. These operations will be referred to as "texture processing". As such, texture processing is best considered at the algorithmic level of Marr's scheme. That is not to say that it does not have goals which constrain its operation, but rather that the goals of texture processing are concerned with delivering higher level visual operations with an appropriate description of the input. This goal may be stated specifically.

Assumption 2 *Texture processing produces a statistical description of an image from a set of spatially distributed texture elements (texels).*

This assumption has two components: existence of a distinct representation for texture information, and the calculation of texture statistics over space. These will be considered in turn.

It is hard to justify the existence of a particular representation within an information processing system. For example, Watt (1988) points out that one can always ask "why

have a primal sketch, why not go straight to the 2.5 D sketch?". He presents two reasons for having an intermediate representation: it can remove information that is redundant to higher levels of processing, and a it can simplify, in a computational sense, algorithms at higher levels. Similarly, it is clear that the alternative to the use of local texel statistics, i.e. the direct usage of individual texel attributes, contains vast redundancy for texture tasks (e..g shape from texture). The level of reliability of individual measures is too low to make their use worthwhile. It is also clear that a system operating on relatively coarse spatial estimates of statistics will be a great deal simpler than one using detailed information from a potentially large number of texture elements.

This assumption that statistical measurement of texel attributes is the basis of texture processing, suggests two stages of processing: measurement of texels and their integration. Is it reasonable to assume that these two operations are occurring in all processing of visual texture? It is in the nature of visual information that it is noisy. Because of irregularities in visual reception and neural noise, reliable visual information requires averaging of some sort. This is evident in the spatial summation properties of cells in V1, etc. To get the most from higher order visual information, such as orientation, it would make sense to perform some similar form of integration.

The specific evidence for *spatial* summation of texture attributes comes mainly from the work of Sagi (Sagi and Julesz, 1987; Sagi, 1990), and of Nothdurft (Nothdurft, 1985b; Nothdurft, 1991b). Sagi and co-workers (Sagi and Julesz, 1987; Sagi, 1990) have examined the identification of a target element, differing in orientation from a background field of distractors, as a function of the number of background distractors, and have found a non-monotonic dependence of performance on distractor density. Performance first decreases as the number of distractors increases, but then begins to improve with very dense numbers of distractors. At short stimulus presentation times, the initial decrease in performance as distractor density increases is attributable to an decrease in the probability of attending to the target. However the increase shown at high distractor densities is evidence for a mechanism detecting feature-differences through local interactions (Sagi, 1990). Similarly, Nothdurft (1985b) has shown that line length interacts with the spacing of texture elements, and that texture boundary detection depends on the rate of change of element orientation in space, at the boundary. This is taken as evidence for the calculation of *gradients* across space (Nothdurft, 1985b). Nothdurft (1991a) has also demonstrated that texture boundaries defined by orientation are resistant to local luminance fluctuation,

which is evidence for a stage of integration beyond V1, where cells are sensitive to image contrast and orientation (Watt, 1991a).

Consider the stimuli that have been devised for psychophysical texture experiments in the past (e.g. Figure 1.2). It is a defining characteristic of every artificial texture task that it cannot be performed by monitoring a single element of the display. This is a weak constraint; many experiments would seem to fulfill this criteria and are not presently considered "texture tasks". Also because of the lack of variation of feature attributes in many stimuli, often only two elements *need* to be monitored to perform the task. The point is that a *difference* or multi-variate function of some sort must be used to do the task.

There is little or no work examining *directly* the limits on human estimation of textural statistics. This requires a different type of psychophysical procedure to typical texture experiments: subjects are asked to perform tasks which require the direct use of features of this statistical description. Watt (1991b) briefly presents results from a task requiring subjects to estimate the mean orientation of a patch of texture elements, where the orientation of texels varies around some mean value. Data indicate that the mean orientation of a number of orientation measures is available to subjects. Chapter 3 compares human estimation of texture statistics, to a set of alternative models for coding average orientation.

Assumption 3 *Orientation is a dimension used to form these descriptions.*

The importance of orientation information to visual processing is demonstrated by the fact that it, along with size, are the first visual dimensions to be represented in the human visual system (e.g. Hubel and Wiesel, 1967). The primacy of orientation as a cue for texture processing has been demonstrated in a number of ways. Beck and his colleagues (Beck, 1966; Beck, 1967; Beck and Ambler, 1972; Beck and Ambler, 1973) have consistently demonstrated that line orientation influences texture segregation more strongly that the "figural similarity" of elements. For example, the patch composed of slanted "T"s in Figure 1.2 is more easily segregated from the patch of upright "T"s than a patch of (figurally different) "L"s.

Nothdurft has established the importance of orientation as a cue for texture segmentation using a number of paradigms. For textures composed of short line segments, discrimination of the shape of patch of elements differing in orientation from the background, improves as a function of the orientation difference between patch and background

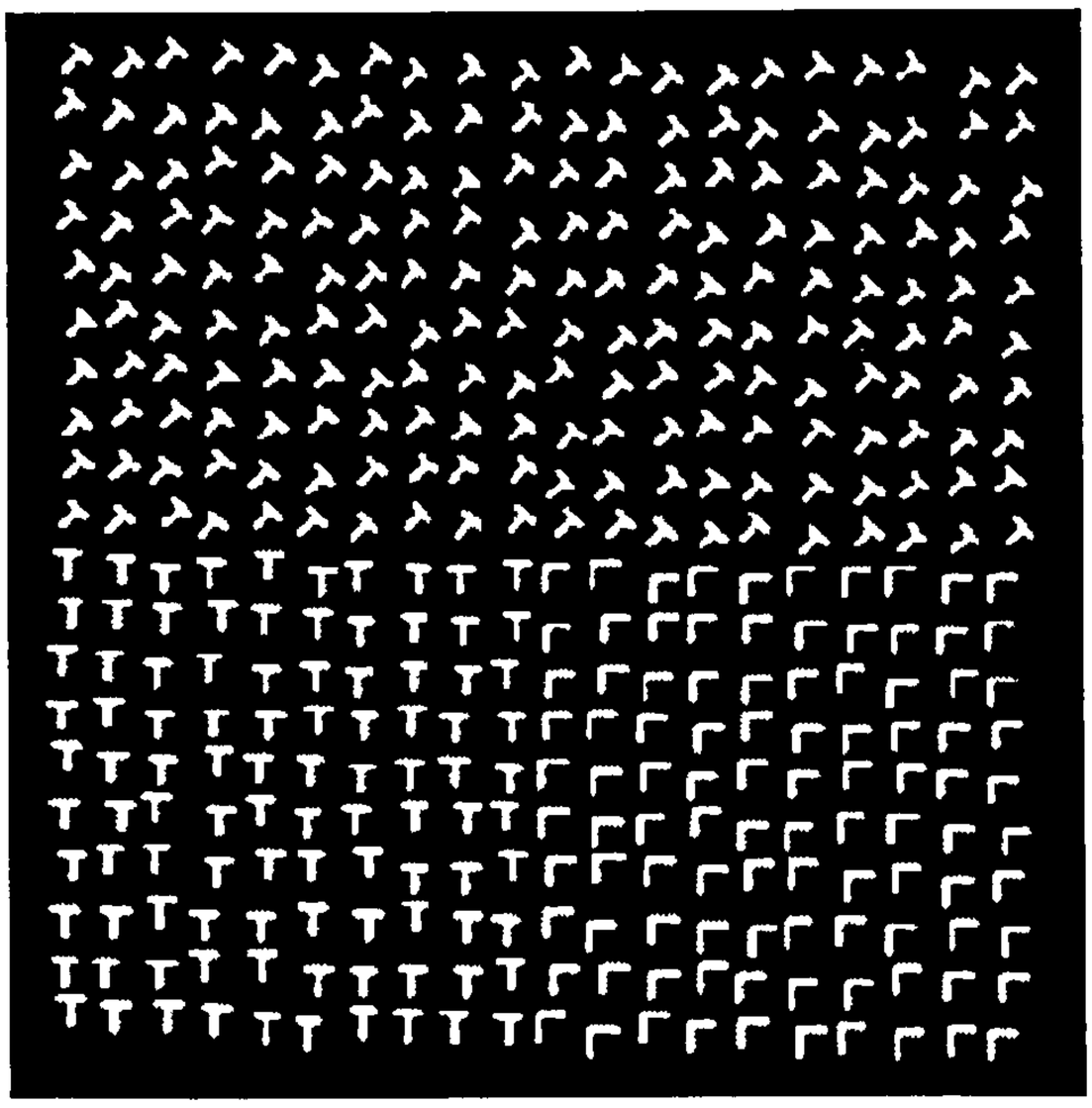


Figure 1.2. Beck's demonstration of the "primacy of orientation".

(Nothdurft, 1985b). This result has been confirmed by Landy and Bergen (1991), using spatially bandpass textures. Nothdurft (1991b) has shown that segmentation using a number of proposed texture attributes is not possible in the presence of local luminance fluctuation. Only orientation is resistant to this form of noise. Nothdurft (1991a) tested the hypothesis that the detection of texture elements within patterns, and the discrimination of two regions composed of different texture elements, should be affected similarly by masking with band-limited visual noise. Results showed that the only supposed texture attribute for which this is true, is orientation.

From the distortions in the geometry of various proposed texture attributes that arise from the shape of a textured surface, psychophysical evidence indicates that it is orientation information that dominates subjects' perception of surface shape (Cutting and Millard, 1984; Todd and Akerstrom, 1987; Blake *et al.*, 1993; Cumming *et al.*, 1993).

Assumption 4 *Texture processing operates over a range of spatial scales, and a single scale is selected to perform a task.*

That texture processing occurs at a range of scales must be true to provide the degree of *scale-invariance* which we show in texture perception. For example, Nothdurft (1985b) has shown that increased spacing of texture elements impairs texture discrimination, only if elements are not scaled in proportion to their spacing. In addition, if texture

information does occur at a range of scales in natural images, as some definitions in the previous section suggested, then this assumption also seems justified. If we did not separate spatial scales and choose between them, then information from independent sources of texture information would always be amalgamated.

This assumption could be restated as: some scales are better than others for performing texture tasks. Spatially band-limited masking studies have certainly demonstrated that this is the case for texture segmentation (Nothdurft, 1991a). Further evidence for the operation of texture processing at single spatial scales, is presented in Chapter 5.

How is scale selected? Ideally to select the correct scale one would perform the task at all scales and check which scale achieved the goal. Given that this can't be done without prior knowledge of the goal (in which case there is no point doing the task!), and that scale must be selected, other (more general) strategies must be employed to select scale.

Assumption 5 *Texture processing makes assumptions about statistical properties and selects scales whose statistics deviate from some expected value.*

Statistical properties are available directly from the texture processing stage, and it is sensible to use those statistics to make the decision as to which description should be selected. In order to make that decision, assumptions have to made about the statistics of images that are inappropriate to texture processing.

For example, the use of *texture gradients*, the increase in density of texture elements due to perspective projection, in deriving surface shape, assumes homogeneity of texel positions. In the case of orientation, isotropy is the assumed state of the world. This assumption is again commonly made in the literature on shape from texture: deviations from isotropy code local surface shape. Witkin (1981) claims that the assumption of texture element isotropy is not a strong, and therefore restrictive, assumption about texture regularity but merely the expectation that texture does not "conspire to mimic projective effects, or to cancel those effects" (Witkin, 1981). Here we are stating that if one assumes isotropy, then the best scale to operate at is maximally anisotropic.

The hypothesis that anisotropy is a useful criterion for selecting scale, is tested in Chapter 5.

Notice that by defining texture processing, but not texture, no constraint has been put on what an appropriate input for the system is. Any image can be treated as texture and it is the contention of this thesis that indeed all images are treated as texture in the

early stages of visual processing. For example, it has recently been suggested (Emmott, 1994) that statistical regularities in symbolic descriptions of the output of LoG filtered pages of text may be used to select the correct spatial scale for reading sub-tasks. This suggests that the visual system may have a whole package of criteria for selecting spatial scale according to the task at hand.

But what of the commonly held notion of texture? When one looks at a set of images one can be fairly confident in identifying what is and what isn't texture. What are these images?

Assumption 6 *"Texture" is a class of visual image which is wholely characterised by its statistical description, in the context of a particular texture task.*

A "good" texture is a stimulus which is deemed, in some sense, *appropriate* to a particular texture operation. Based on the original definition of the goal of texture processing, (Assumption 1), appropriateness will depend upon the presence of multiple, spatially distributed elements with measurable attributes. This assumption is in accord with definitions of texture which make reference to the scale of the texture compared to the surface e.g. "fine structure", "detailed structure too fine to be resolved", etc. Fine scale implies a great number of individual measures, for which a statistical characterisation is appropriate.

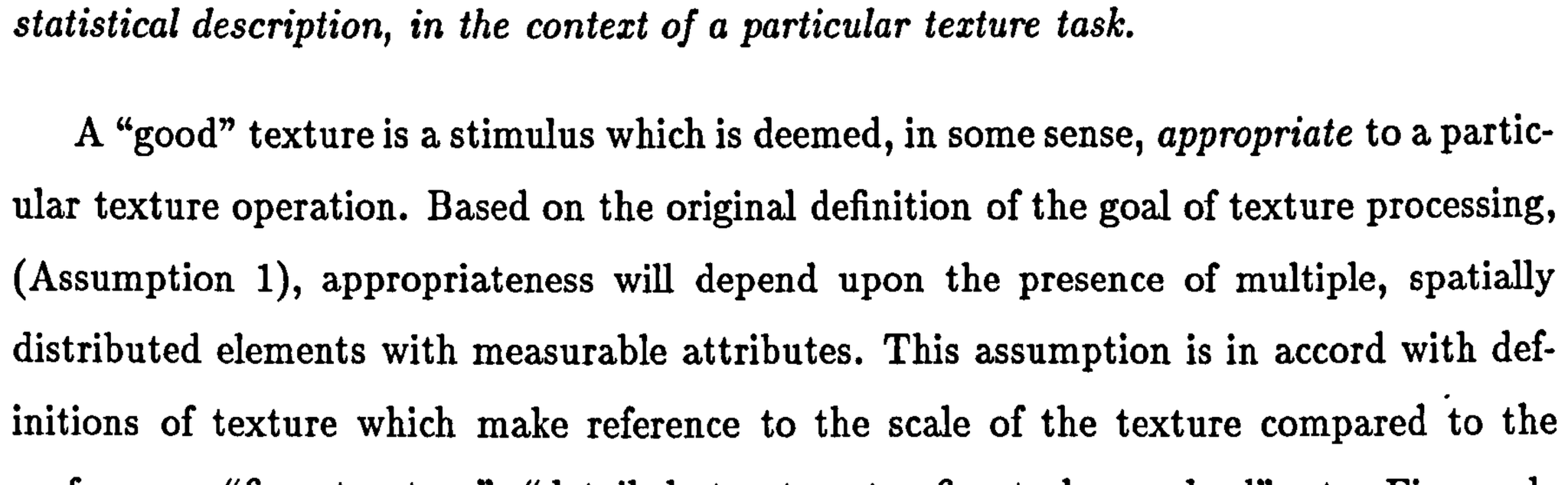


Figure 1.3. (a) Enlarged section of face image, (b). (c) Reduced version of the face image, repeated and embedded within copies of another face.

For example consider Figure 1.3. Of this set of images, parts (a) and (c) would subjectively be considered texture, and part (b), an object. Figure 1.3a shows a magnified region of (b), and (c) shows a number of reduced versions of (b) embedded in a background

of other faces. Because both (a) and (c) do not contain readily recognisable faces, but instead, are well described by their statistical representations, they are termed texture. Of course these judgements are made without the viewer being given a particular visual task to perform on the image. Since Figure 1.3b, in addition to the possibility of being described statistically, is also appropriate to higher recognition processes, this collection of visual features is treated as a face. This further illustrates the way in which scale selection is intimately related to the task at hand. By forcing the viewer to use particular scales, different tasks become appropriate.

1.4 Present research into texture

In this section a brief overview is given of research into three visual operations that use texture information: shape from texture, texture segmentation, and texture flow.

1.4.1 Shape from texture

Shape from texture is a set of techniques for recovering the three-dimensional structure of a surface from the distortions in texture which occur when a receding surface is projected onto the image plane. Three types of cue are generally acknowledged to arise these distortions: perspective, density and compression (Cutting and Millard, 1984).

Perspective cues refer to the inverse scaling of texture elements that occurs as the surface they cover recedes from us. However, statistical modelling of this process is not possible without prior knowledge of the texel size statistics of a surface at a known orientation. For that reason, perspective cues are the least investigated of the cues to surface shape, and are generally eliminated from experimental stimuli either by restricting the field of view of a stimulus (which approximates orthographic projection), or through the use of thin texture elements (e.g. lines) on which perspective scaling will have a negligible effect (Blake *et al.*, 1993).

Gibson (1950b) first proposed that, assuming surfaces are uniformly covered with markings, the change in density of texture markings, under perspective projection, indicates the orientation of the surface. This "texture gradient" can thus be used to make multiple, local estimates of surface orientation and so overall shape. Several models have subsequently been proposed using and extending this idea (Bajczy and Lieberman, 1976; Ikeuchi, 1984; Aloimonos and Swain, 1984; Kanatani, 1984). However for the density cue to be useful, it is necessary to assume homogeneity of texel positions on a flat surface. As

Witkin (1981) points out, the homogeneity assumption has never been tested. Furthermore, only density *gradients*, and not the absolute degree of density, can give information about shape (Blake *et al.*, 1993).

As a surface recedes away from us, the markings that cover it appear compressed in the direction that the surface recedes (*foreshortening*). Under projection, the distribution of texture element orientations will be biased towards one that is orthogonal to the direction of surface recession. In order to make use of this *compression cue*, Witkin (1981) makes the assumption that *texture orientation is initially isotropic*. Note that in contrast to the density cue, the absolute degree of compression does give information about shape. Although, implicitly, comparison to the statistics of a true fronto-parallel surface is made, a gradient within the image is not necessary.

There is also increasing psychophysical evidence that it is the compression cue that is critical to subjects' perception of surface shape (Cutting and Millard, 1984; Todd and Akerstrom, 1987; Blake *et al.*, 1993; Cumming *et al.*, 1993). For example, Cumming *et al.* (1993) covaried the three cues discussed above, using stereoscopically viewed, artificially textured surfaces. Using a two-alternative judgement of depth, 97% of variation in subjects' results is accounted for by the compression gradient. Blake *et al.* (1993) have developed an ideal observer model for shape from texture, using density and compression cues, in order to predict the lower bounds on performance at shape judgements. They have shown, without recourse to the cue conflict paradigms of earlier work (e.g. Cutting and Millard, 1984; Todd and Akerstrom, 1987), that human shape judgements exceed the performance of a model using density cues alone and that other cues, such as compression, *must* be being used.

A number of computational models have been proposed which make use of the compression gradient, in the form of the deviation of texel orientation from isotropy, to derive shape from texture (Witkin, 1981; Davis *et al.*, 1983; Kanatani, 1984; Blake and Marinos, 1990). An early and highly influential model, of Witkin (1981), is described in more depth in the following chapter.

1.4.2 Texture segmentation

Texture segmentation refers to the inference of physical surface discontinuities from sudden changes in the measurement of particular image attributes, such as orientation, intensity, size, etc. (Marr, 1982) However, it is unusual for objects in natural scenes to be

distinguished from their backgrounds purely by differences in the texture covering them. Colour and brightness, for example, will often serve to segregate surfaces unambiguously, and our reliance on these cues is reflected in our excellent acuity for locating edges defined along those dimensions. Texture information, however, will be resistant to exactly the type of noise that disrupts segregation based on other more generally reliable perceptual dimensions. For example, variations in local luminance due to shadow will not affect texture boundaries. For that reason segmentation based on texture is an important aspect of visual processing.

Initial interest in the study of texture perception was motivated by the phenomenon of texture segmentation. Julesz (1962) used random-dot texture pairs to show that segmentation of such textures occurs if the intensity of component dots, or if the average distance between dots is significantly different. However, manipulations of the statistics of dot triples does not produce segmentation. This led Julesz to propose that only the first and second order statistics of textures are used for segmentation (Julesz, 1962; Julesz *et al.*, 1973). Experimental results have shown this view to be over-simplistic (e.g. Enns, 1986; Beck *et al.*, 1987; Nothdurft, 1990; Gurnsey, 1987). Nothdurft (1985b) critically showed that it was *local* differences between texture elements that determined discriminabilty.

Although Julesz's theory expanded to accommodate locally conspicuous image features, it was never specified in sufficient depth to be physically realised. Instead it was found that many segmentation phenomena could be explained in terms of spatial filtering (Harvey and Gervais, 1978; Harvey and Gervais, 1981), and consequently a large number of models were proposed to perform texture segmentation using these methods (e.g. Caelli, 1985; Coggins and Jain, 1985; Beck *et al.*, 1987; Sutter *et al.*, 1988; Fogel and Sagi, 1989; Bergen and Adelson, 1988; Bovik *et al.*, 1990; Malik and Perona, 1990). Without going into great detail there seem to be three necessary components of all these models (Chubb and Landy, 1991):

- Convolve the input image with linear filters at various spatial scales. Typically directional derivatives are extracted using various methods, such as: DoGs (e.g. Bergen and Landy, 1991), Gabors (e.g. Malik and Perona, 1990), derivatives of Gaussians (e.g. Bergen and Adelson, 1988), etc.

- Apply a non-linear function of some sort. Examples include, half-wave rectification (e.g. Malik and Perona, 1990), full-wave rectification (e.g. Bergen and Adelson,

1988), or energy computation (the output of filters in quadrature phase are squared and added, e.g. Fogel and Sagi, 1989; Sutter *et al.*, 1988; Bovik *et al.*, 1990).

- Partition fields according to how consistent the derived output functions are. This is usually a straightforward edge detection operation on the output of the previous stage.

All of these models perform segmentation adequately, and even claim a degree of biological plausibility. They also all tend to suffer similar problems:

- Setting of critical variables, such as the spatial scale of filters, by hand.

- Over-specificity. Most models work only for segmentation and, given the apparent scarcity of pure texture boundaries in natural scenes, it seems unlikely that human vision dedicates such complex systems to detect them.

- Poor evaluation of performance: subjective evaluation of segmentation performance tends to be used. Any comparison to psychophysical data often uses rank order agreement of strength of segmentation, of various textures, for human and model.

1.4.3 Texture flow

Consider the image shown in Figures 1.1c. Why should the visual system be able to perceive "static flow" in images such as water or fur? As an image statistic, local flow information is a form of regularity which allows for an economical representation of texture and may also give information about occlusion and shape. In addition, flowing textures are the result of many different natural processes (Kass and Witkin, 1985) and so also offer insight into the formative processes a surface may have undergone (e.g. the stretching which forms woodgrain). This is useful if one wishes to derive rigidity information about a surface and highlights the way in which texture operations seek to *explain* rather than *describe* images (Witkin and Tennenbaum, 1983).

Kass and Witkin (1985) have proposed that the detection and description of image flow is a basic perceptual primitive. The recovery of the physical processes which produce such image flow is viewed as an essential stage in decomposing an image into its intrinsic parts. They consider three of these physical processes which lead to oriented image structure: propagation (e.g. paint streaks), accretion (e.g. rock strata), and deformation (e.g. stretched fabric). The system they describe operates by deriving the flow field, and

deriving the residual pattern that remains when the image is described using a co-ordinate system built from the flow-field. Local flow direction is taken to be perpendicular to the direction in which most zero-crossings lie. Such a system is applied to a number of tasks involving flow, such as the detection of orientation singularities, and anomalies.

Zucker (1982) has termed the task of inferring a set of tangents to the visual contours in the image, "orientation selection". The most popular representation for this information is a *vector field*, which describes the local tangent direction at every point in the image. Zucker (1982) has proposed that such a level of description, which falls between raw image and a fully grouped representation, operates in human vision. He also proposes that there are two fundamental structural classes of orientation fields: Type I and Type II. Type I fields are composed of dense, well defined contours, whereas Type II fields are a result of highlight and reflectance change, and are much more sparsely defined. In his scheme, local flow direction is computed by convolving the image with a DoG filter and using the orientation of the most active filter at a point. These local measurements are integrated into coherent contours by a process of iterative relaxation, which attempts to maximise local co- circularity, the consistency of orientation measures with the hypothesis that they result from a continuous contour.

1.5 Structure of the thesis

In summary, this thesis considers two questions:

- What is an appropriate representation for oriented visual textures, and what computational mechanisms could be used to derive it?
- Do humans construct such a representation and if so, how?

Chapter 3 looks at the representation of *statistical* information derived from texture. Chapters 4 and 5 examine the selection of spatial scale and the extraction of texture elements from texture. Chapter 6 consider the way in which global orientation structure is preserved in the model.

The structure of this thesis is then as follows:

- Chapter 1considers the problem of defining visual texture, and proposes a statistical representation for information derived from texture processing. Research on three tasks that use texture information is introduced.

- Chapter 2 describes an algorithm for deriving this representation from images (the *adaptive filtering model*), and considers its application to four visual tasks.

- Chapter 3 presents psychophysical and simulation data, relating to the way texture statistics are computed and represented. Specifically it tests one component of the algorithm proposed: that representative local orientation is computed using the centroid of a number of orientation measures.

- Chapter 4 examines the way in which scale is selected for texture processing, using psychophysical tasks which require precise scale selection to extract texture elements (Glass patterns).

- Chapter 5 compares the predictions made by three models for deriving texture elements, of performance on the task presented in Chapter 4.

- Chapter 6 describes three experiments looking at the preservation of *global* orientation structure, i.e. systematic orientation change in the image, by the adaptive filtering model.

- Chapter 7 looks at the effect of local contrast variation on deriving texture orientation information.

- Chapter 8 relates findings to the view of texture described in Chapters 1 and 2.

2 | Representing texture flow

2.1 Overview

This chapter presents a model for computing and representing local orientation statistics from visual texture. The model has 4 stages:

- Convolution with Difference-of-Gaussian filters at different spatial scales and orientations.
- Selection of filter responses across orientation: at each spatial scale a point by point selection is made of the output of the filter with the greatest "power" (i.e. squared output). This result is termed the "adaptively filtered" image.
- Extraction of texture elements using half-wave rectification and symbolic description of "blobs".
- Calculation of texel orientation statistics including a measure of mean local orientation variance over the entire pattern.

The model yields a set of texel statistics at each spatial scale: to select a single scale the mean local orientation variance is minimised across the texel sets. The final representation is a set of local estimates of mean orientation and orientation variance. It is proposed that this information is useful for a number of tasks that may require integration of orientation estimates. Four examples of such tasks are considered in the context of the model:

- Texture segmentation.
- Image flow.
- Shape from texture.
- Contour integration.

2.2 INTRODUCTION

This chapter proposes a model for the computation and representation of orientation statistics from visual texture. The rationale behind the approach has been described in Chapter 1. The key differences between this and other models of the representation of orientation are the use of texture elements rather than filter energy and the automatic selection of spatial scale. These two issues will be discussed briefly in this section.

2.2.1 TEXTURE ELEMENTS VERSUS FILTER ENERGY

Early theories of texture perception (Julesz, 1981; Julesz, 1986; Beck, 1972; Beck, 1982) proposed that texture discrimination occurs by means of differences in first-order statistics of the attributes of a set of image features. Because texton theory typically employed black and white textures the problem of how to derive texture elements automatically from natural images has largely been neglected. The candidates proposed are typically features in the output of isotropic filters, such as zero-crossings in the output of Laplacian-of-Gaussian ($\nabla^2 G$) filters (Witkin, 1981; Wen and Fryer, 1991), descriptions of blobs from half-wave rectified output of $\nabla^2 G$s (e.g. Voorhees and Poggio, 1987), "disks" matched to $\nabla^2 G$ responses (Tomita *et al.*, 1987), etc.

However a number of recent models, particularly of texture segmentation, do not attempt to derive explicit image features from filter outputs but instead use directly the local filter energy as an indication of texture statistics (Malik and Perona, 1990; Bovik *et al.*, 1990; Fogel and Sagi, 1989). This is adequate if assuming that the information derived from texture processing will not be used by any further visual processes. All that a filter energy model for segmentation provides is an indication that a certain part of the image differs significantly from another with respect to its local texture statistics.

Texture elements are clearly necessary for a representation which is to be used for further processing because such processing involves associating attributes with local image features. Take the extraction of surface contours as an example. This involves the isolation of components of contours and some form of integration process (e.g. theta aggregation, Marr, 1976). Texture elements are a natural input to this process but there is no obvious way that filter energy information, in isolation from detailed spatial information, can be used.

2.2.2 Filter selection

The performance of any model employing spatial filters will be affected by the scale of filter used. However the issue of how one sets filter size *automatically* has rarely, if ever, been considered. Filter size is usually set by hand to give the best results (e.g. Bovik *et al.*, 1990; Malik and Perona, 1990; Voorhees and Poggio, 1987). This is entirely due to the poor specification of the task. Given the output of the model at number of spatial scales, how would one determine the degree of successful segmentation? A clear task definition would allow for the formulation of performance measures which could be simply maximised across scale to give optimal performance of the model.

In this chapter it is proposed that the statistics of texel attributes provide a simple criterion for selecting the right spatial scale to use. Specifically a measure of local orientation variation is calculated and minimised to select the correct spatial scale. It will be shown that this is a useful criterion for a number of tasks operating on the output of the model.

The structure of this chapter is as follows: in the next section the preprocessing stage of the model is presented - the details of how texture elements are extracted from the image. The next section describes how local orientation statistics of those texels are computed and the application of those statistics to the automatic selection of filter scale. Four applications of the model, to tasks requiring the use of texture orientation, are described in the final section.

2.3 Preprocessing stage of the model

The goal of the initial stage of processing is to produce a symbolic description of the texture elements contained in an image, at each spatial scale. A schematic view of this stage of the model is shown in Figure 2.1.

The model uses a bank of Difference-of-Gaussian filters, operating at S spatial scales and T orientations. Filters are denoted $F_{\sigma,\theta}(x,y)$, where σ and θ index the filter scale and orientation respectively. Filters are convolved with an input image, $I(x,y)$:

$$R_{\sigma,\theta}(x,y) = F_{\sigma,\theta}(x,y) * I(x,y)$$

This representation is squared to give, $R^2_{\sigma,\theta}(x,y)$, an estimate of the power at each spatial scale and orientation. This is smoothed by convolving it with a Gaussian filter, G_σ:

$$GR^2_{\sigma,\theta}(x,y) = G_{2\sigma} * R^2_{\sigma,\theta}(x,y)$$

where $G_{2\sigma}$ is a Gaussian filter with a standard deviation equal to twice the standard deviation of the DoG filter at that scale:

$$G_f(x,y) = \frac{1}{\sqrt{2\pi}f} e^{\frac{-(x^2+y^2)}{2f^2}}$$

The next stage selects the pixel from the output of the most active filter at each point:

$$A_\sigma(x,y) = R_{\sigma,\theta}(x,y) : \max_{\theta=0,1,...T}[GR^2_{\sigma,\theta}(x,y)]$$

Texels are then extracted from this "adaptively filtered" representation. This is done by half-wave rectification of the output at grey levels above and below a value equal to the standard deviation of the image grey-levels. If $A^+(x,y)$ and $A^-(x,y)$ indicate the positive and negative portions of the adaptively filtered image, respectively then:

$$A^+_\sigma(x,y) = \max[A_\sigma(x,y), \tau]$$

$$A^-_\sigma(x,y) = \min[A_\sigma(x,y), -\tau]$$

where, for an image with dimensions X by Y:

$$\tau = \sqrt{\frac{\sum_x \sum_y (A(x,y) - \mu)^2}{XY}} \tag{2.1}$$

and

$$\mu = \frac{\sum_x \sum_y A(x,y)}{XY}$$

$\sum_x \sum_y$ denotes summation over all image pixels.

Each of those blobs is then characterised using an image description scheme (Watt, 1991b). This yields a set of "sentences" describing each blob in terms of its position, orientation of principal axis, length, etc. For the ith blob in a set of N total blobs its orientation, θ_i, and mass, M_i are used for calculating orientation statistics. Such a description is generated at each spatial scale and all of the texel sets are passed on to the next stage of processing.

The following subsections describe the components of the model in more detail and also consider the rationale behind their choice.

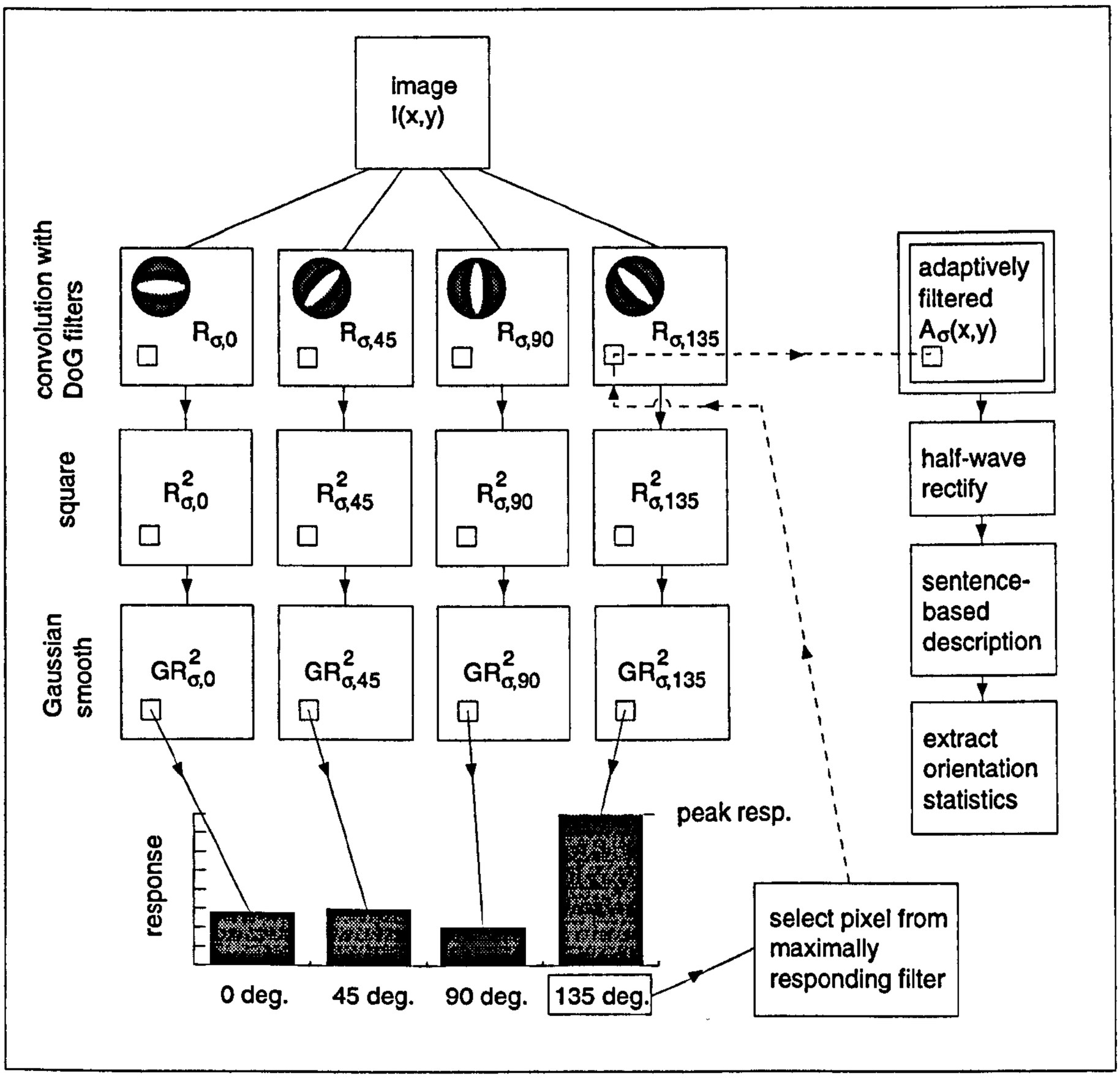


Figure 2.1. Overview of the adaptive filtering model operating at one spatial scale, σ. The image (top) is convolved with DoG filters to give $R_{\sigma,\theta}(x,y)$. These images are squared, and smoothed with a Gaussian filter to give $GR^2_{\sigma,\theta}(x,y)$. This representation is used to select, point-by-point, the grey-level associated with the maximally responding filter. This grey-level is copied into the adaptively filtered image and this image is used to extract texels.

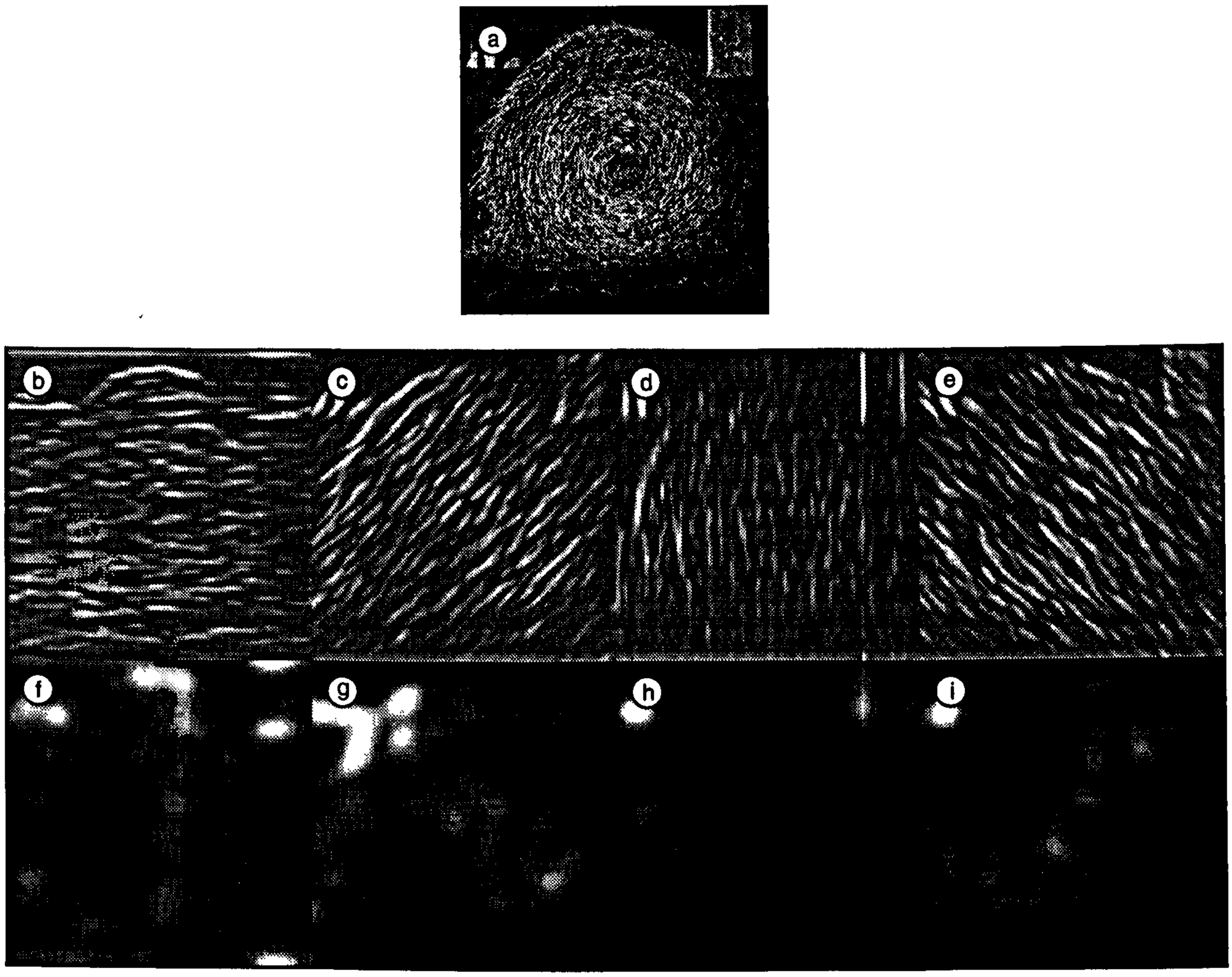


Figure 2.2. Operation of the adaptive filtering model, at one spatial scale, on an image containing rotational flow field. (a) Original image. (b-e) Output of oriented DoG filters. (f-i) Squared output of DoGs. (j) Output of maximally responding filter at each point and (k) texels derived using half-wave rectification.

2.3.1 Spatial filtering

A variety of models have been proposed as descriptions of the oriented, point-spread function of cells in primary visual cortex (V1) including Gabors (Daugman, 1980; Daugman, 1985), Difference-of-Gaussians (Wilson and Gelb, 1984), and Difference-of-offset-difference-of-Gaussians (Parker and Hawken, 1988; Malik and Perona, 1990). There is no evidence to suggest that the differences between these models are critical; the two-dimensional Difference-of-Gaussians (DoG) was chosen because it is computationally simple and its characteristics have been determined using a number of psychophysical techniques (e.g. adaptation, Wilson and Regan, 1984; sub-threshold summation, Mostafavi and Sakrison, 1976; masking, Wilson *et al.*, 1983). These filters are composed of a DoG in the x-direction multiplied by a Gaussian in the y-direction:

$$F_{\sigma,\theta}(x_t, y_t) = \left(e^{-x_t^2/2\sigma^2} - (1/2.23)e^{-x_t^2/2(2.23\sigma)^2}\right) e^{-y_t^2/2(3\sigma)^2} \tag{2.2}$$

where σ refers to the standard deviation of the positive Gaussian function. x_t and y_t are coordinates rotated by angle ϕ:

$$x_t = x\cos\phi + y\sin\phi$$

$$y_t = y\cos\phi - x\sin\phi$$

The ratio of the amplitudes of the positive and negative parts of the DoG and the aspect ratio of the filter are based on estimates derived by Wilson and co-workers (Wilson and Gelb, 1984; Phillips and Wilson, 1983;).

Note that this is an even-symmetrical filter although it is known that receptive field profiles of cells in V1 vary continuously between odd and even symmetry (Parker and Hawken, 1988). However, human texture processing does not seem to be able to make effective use of odd-phase information. Textures composed of odd-phase Gabor patches, where one half of the pattern is the mirror-reverse of the other, do not produce preattentive segmentation (Rentschler *et al.*, 1988; Malik and Perona, 1990). To segment these textures successfully requires the use of odd-phase filters, which suggests that such mechanisms are not relied upon by the visual texture processing system. For this reason phase information is not preserved by the model.

The image, I, is convolved with a bank of 84 DoG filters, $(F_{\sigma,\theta})$. The filter bank contains 12 orientations and 7 spatial scales of filter ($T = 12$, $S = 7$). The standard devi-

ations of the DoGs are 1-16 pixels in half octave steps, corresponding to peak sensitivities of 40.7-2.5 cycles per image. Their orientations varied between 0° – 135° in 15° steps.

Figure 2.2 shows the result of filtering a typical oriented texture, the haystack image (a), with filters at one spatial scale and four orientations: (b) 0°, (c) 45°, (d) 90° and (e) 135°. Note that the filter outputs are particularly high around the outer contour of the haystack and that the vertical filter responds strongly to the feature in the top left hand corner of the image.

Given these outputs at different orientations, the question is, how does one *combine* their outputs to reflect local orientation accurately?

2.3.2 Combining filter responses

A simple combination rule for the filter outputs is to add the outputs together, so that filters which are locally responding more strongly will dominate that region of the output image. The problem with this method is that by collapsing across orientation, by summing oriented filter output, one has effectively simulated the output of an isotropic filter (see Figure 2.4 for a comparison of the output of a $\nabla^2 G$ and summed DoGs). What is required is a mechanism which locally suppresses weak filter responses and only uses the output of the most strongly responding filter mechanisms.

Inhibition between cells in V1 with similar spatial frequency sensitivity and between cells with similar orientation sensitivity has been observed (Bishop *et al.*, 1973; Burr *et al.*, 1981; Morrone *et al.*, 1982; Morrone and Burr, 1986; Tsumoto *et al.*, 1979), but the role of inhibition remains unclear. It has been proposed that the function of (intracortical) inhibition is the suppression of spurious responses from nonoptimally tuned channels (Paradiso, 1988; Malik and Perona, 1990). The reduction of a cell's orientation bandwidth, which results from intra-orientation inhibition, could not only serve to increase orientation discrimination (Paradiso, 1988), but could specifically enhance a cells response to contours (Morrone *et al.*, 1982).

Malik and Perona (1990) have realised this inhibition computationally, as a part of their model of texture segregation. The outputs of filters at different orientations and spatial scales are suppressed if they fall below a threshold, a so-called "leaders-take-all" system. A simpler "winner-takes-all" system is proposed here, where the most active filter at each point is used. The principal difference between this and Malik and Perona (1990)'s system is that the peak filter output is not chosen across both scale and orientation but

only across orientation. Filter scale is selected separately, using a combination of the image description and the constraints of the task. A degree of Gaussian smoothing is added to the representation of the local filter energy. This is to avoid the effects of noise and also to enhance filter activity along extended contours.

Inspired by the evidence of intra-orientation inhibition between cells an alternative to the pointwise addition of filter responses is proposed: the selection of the output of the *most active filter*. To perform this selection, an intermediate representation is computed, $R^2_{\sigma,\theta}(x,y)$, which is simply the square of the original filter response image. This image contains the unsigned magnitude of a filter response, for each point in the image. The R^2 image is smoothed by convolving with a Gaussian filter, the standard deviation of which is twice the standard deviation of the DoG filter in use, to give $GR^2_{\sigma,\theta}(x,y)$. Finally this representation is used to make a pointwise selection of the output of the most active filter (from $R_{\sigma,\theta}(x,y)$), over the T filter orientations:

$$A_\sigma(x,y) = R_{\sigma,\theta}(x,y) : \max_{\theta=0,1,\ldots T}[GR^2_{\sigma,\theta}(x,y)]$$

The operation of pointwise selection of the maximally active filter is termed *adaptive filtering*. Figure 2.3 shows an example of the preprocessing stage of the model operating on the haystack image. Notice that in this example the blobs in the adaptively filtered images are highly elongated and their orientation reflects the subjective direction of local flow in the texture.

Given the complexity of this process, what are the advantages of adaptive filtering over simpler filtering mechanisms? Laplacian-of-Gaussian filtering is a computationally efficient method for deriving features (e.g. Voorhees and Poggio, 1987; Marr, 1982; Wen and Fryer, 1991); because the filter kernel is isotropic, only one convolution is required at each spatial scale. By half-wave rectifying the image, blobs are produced which may be described *symbolically* in terms of key attributes such as length, brightness, etc. (Voorhees and Poggio, 1987; Watt, 1991b), in the spirit of a "primal sketch" description (Marr, 1976). However, there are two basic problems with $\nabla^2 G$ filtering. Firstly, it is not biologically plausible, given that V1 simple cells are orientationally selective (Hubel and Wiesel, 1967). Secondly, for operations requiring accurate orientation estimates, such as contour analysis or shape-from-texture, $\nabla^2 G$ filter estimates appear to be too noisy. Figure 2.4e shows the haystack image which contains compelling rotational orientation structure. $\nabla^2 G$ filtering and thresholding this image (Figures 2.4f-h) reveals that very few

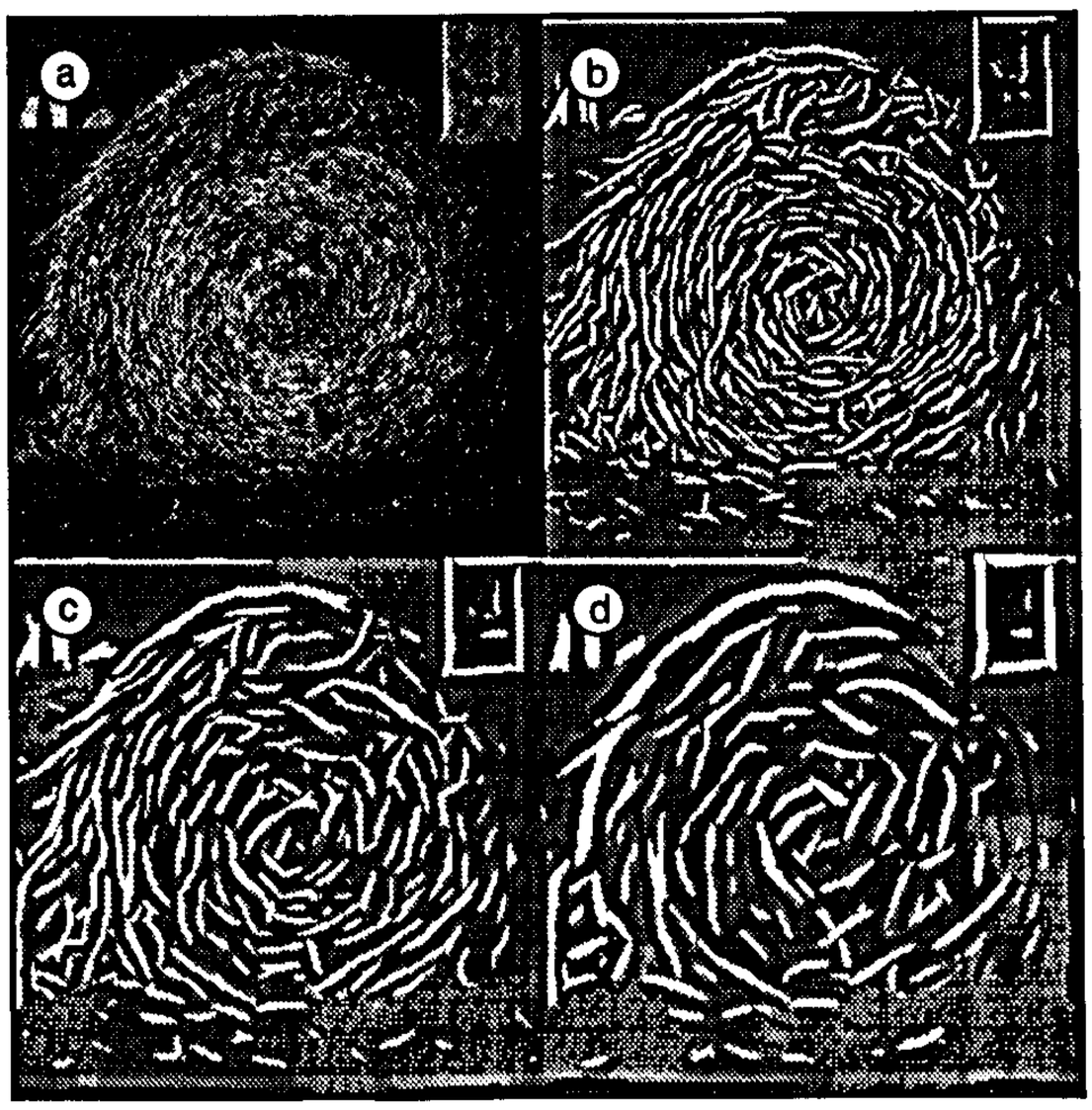


Figure 2.3. Texels derived from the haystack image, (a), using the adaptive filtering model at three spatial scales: (b) 57.6 cycles per image (c.p.i.), (c) 40.7 c.p.i., and (d) 28.8 c.p.i. Grey levels above and below threshold luminance have been replaced by white or black, respectively, in order to enhance reproduction. Note the elongated blobs in (b-d).

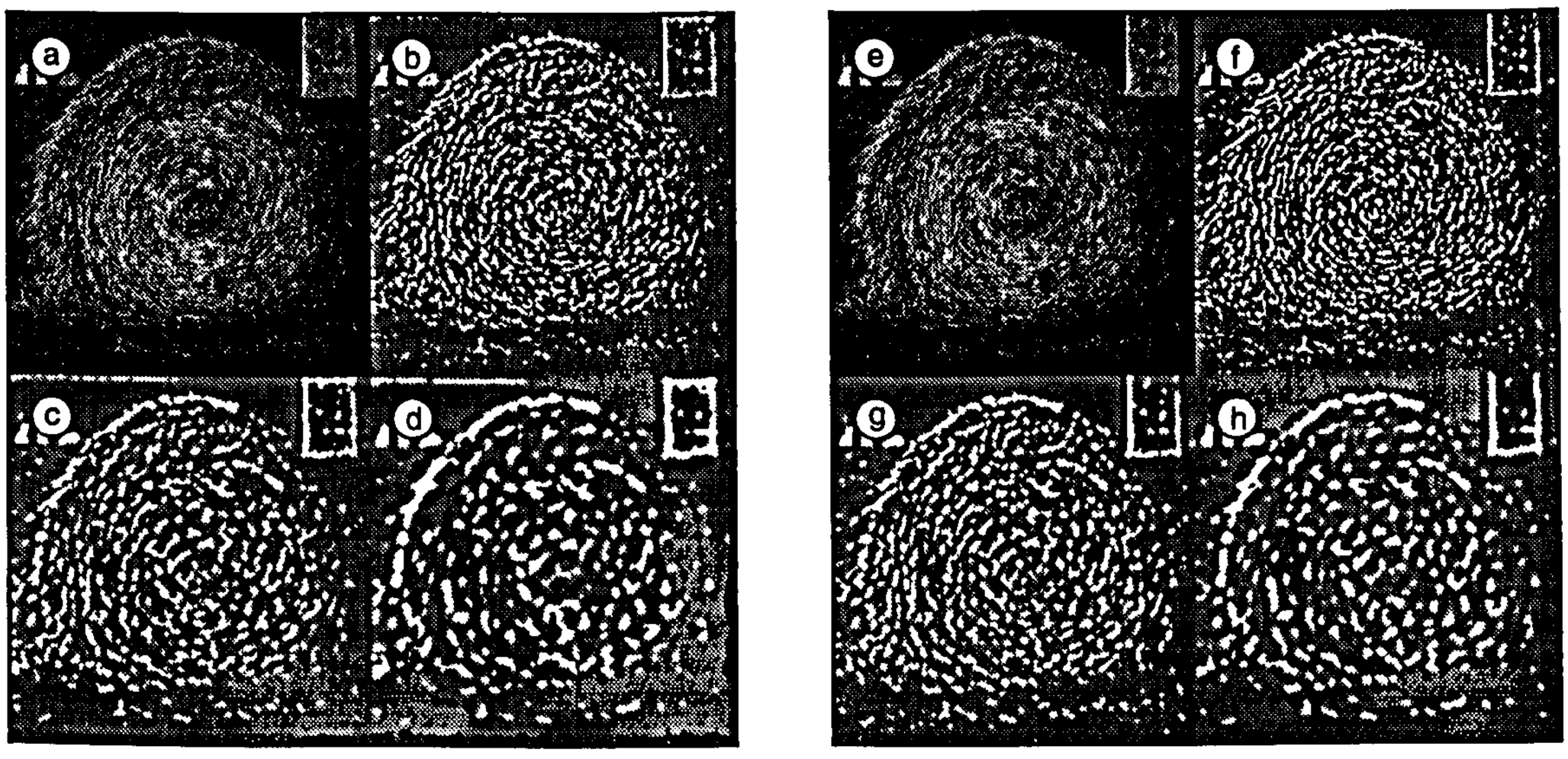


Figure 2.4. (a,e) Haystack image. Texels derived using (b-d) summed DoG outputs and (e-h) $\nabla^2 G$ filtering. Filter output is half-wave rectified: grey levels above and below threshold luminance have been replaced by white or black, respectively, to enhance reproduction. Compare to Figure 2.3 and note that many of the blobs appear to be due to noise and are not oriented in the direction of local image flow.

of the resultant blobs are aligned with the direction of flow. Particularly at fine spatial scales, many blobs are actually unoriented and are probably attributable to noise. These features will produce a very noisy representation of local orientation[1].

2.3.3 Describing the set of texels

The final part of preprocessing is the extraction of texture elements from the adaptively filtered image. By thresholding (*half-wave rectifying*) the image at ± 1 grey level standard deviation (defined in Equation 2.1) "islands" of response are produced in the image (e.g. Figures 2.3b-d). Malik and Perona (1990) point out that the use of half-wave rectification is a desirable form of non-linearity for two reasons. Firstly it is biologically plausible; V1 cortical cells maintain a low discharge rate and cannot signal a negative response through a decrease in firing. Instead two cells are employed to code positive and negative parts of the response - the *on* and *off* pathways, respectively. Secondly, half-wave rectification retains the sign of the signal, whereas full-wave rectification and energy computation both discard it. The signal sign is essential for discrimination between opposite contrast textures.

The model is not highly sensitive to the value of the threshold (τ) at which the image is half-wave rectified. The standard deviation of the image grey levels is used but other measures, such as statistics based on the gradient image (Voorhees and Poggio, 1987), would suffice.

The blobs that result from the rectification stage can be described using procedures to extract their centroids, lengths, etc., and indeed such schemes have previously been proposed as practical methods for deriving texel statistics (e.g. Voorhees and Poggio, 1987, Wen and Fryer, 1991). This model uses an image description scheme developed by Watt (1991b) to generate a sentence-based description of blobs.

Specifically feature descriptions were calculated as follows (summarised from Watt, 1991b): given the nth zero-bounded blob in an image, and the ith pixel within it, the centroid is calculated as:

$$(\mathrm{cx}_n, \mathrm{cy}_n) = \frac{\sum(x_i, y_i)}{\sum v_i}$$

where v_i is the brightness of the ith pixel within a blob, (x_i, y_i) is its location, and $\sum$ denotes summation over all pixels in a zero-bounded region.

[1] A comparison of human psychophysical data (from a texture orientation discrimination task) and predictions from the $\nabla^2 G$ and DoG models appears in Chapter 5.

The blob mass is simply the sum of all pixel brightnesses within it:

$$\text{mass}_n = \sum v_i$$

The following calculations are simplified if pixel locations are defined relative to the blob centroid:

$$(x'_i, y'_i) = (x_i - \text{cx}_n, y_i - \text{cy}_n)$$

The variance of the pixel distribution in a direction, θ, is:

$$\sigma_\theta^2 = \frac{\sum v_i (x'_i \cos\theta + y'_i \sin\theta)^2}{\sum v_i}$$

and the slope of the principal axis of blob n is (by differentiating σ_θ^2 with respect to θ, and equating to zero):

$$\text{ori}_n = \frac{1}{2}\tan^{-1}\left(\frac{2\sum v_i x'_i y'_i}{\sum v_i x'^2_i - \sum v_i y'^2_i}\right)$$

Given this angle, the length and width of the blob are given by σ_θ^2 in directions ori_n and $\text{ori}_n + 90°$, respectively:

$$\text{length}_n = \frac{\sum v_i (x'_i \cos \text{ori}_n + y'_i \sin \text{ori}_n)^2}{\sum v_i}$$

$$\text{width}_n = \frac{\sum v_i (x'_i \cos(\text{ori}_n + 90°) + y'_i \sin(\text{ori}_n + 90°))^2}{\sum v_i}$$

In order to calculate a measure of the usefulness of the orientation estimate, ori_n, we use information about the shape of the blob and its size. Specifically, it is assumed that the contribution of a blob orientation to the calculation of local orientation statistics should be weighted in proportion both to its mass and its aspect ratio. A circular blob should be treated as totally unreliable estimate of local orientation as should a blob with a mass approaching zero. This is incorporated into the model as a measure of orientation reliability:

$$\text{rel}_n = \frac{\text{mass}_n \text{length}_n}{\text{width}_n}$$

Finally we re-subscript the orientation and orientation reliability measures by the centroid of the zero-bounded blob they describe:

$$\theta(x, y) = \text{ori}_k : (x, y) = (cx_k, cy_k)$$

$$\mathrm{Rel}(x,y) = \mathrm{rel}_k : (x,y) = (cx_k, cy_k)$$

Thus the final output of the preprocessing stage is a set of S texel sets, one at each spatial scale. Texel sets contain sentences describing the orientation of the principal axis of each blob and a measure of the reliability of this estimate.

2.4 Estimating texel orientation statistics

This section considers methods for deriving useful orientation statistics from texel descriptions. The first section considers the spatial sampling of the texel sets, while the second and third describe how the mean and standard deviation of orientation statistics are computed.

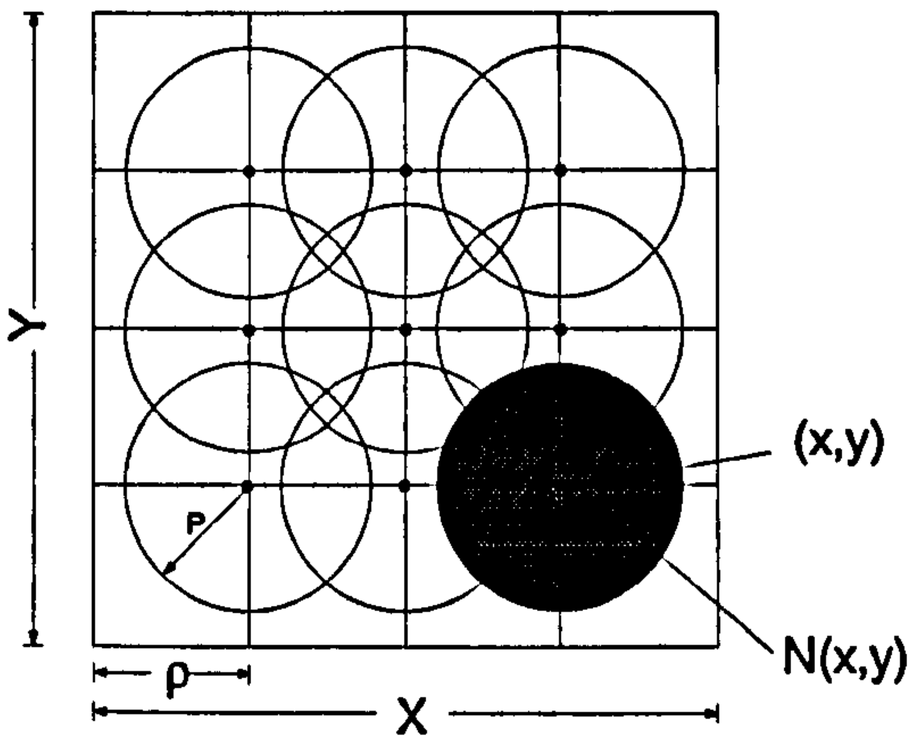


Figure 2.5. Illustration of sampling lattice used in Section 2.4.1

2.4.1 Partitioning the image

Given a number of texel sets derived from filters at scale σ, the following section defines how elements of the sets are spatially sampled to derive local orientation statistics. Figure 2.5 illustrates the sampling lattice used below.

Denote the image lattice (the set of all pixel positions) $\{(i,j) : 1 \leq i,j \leq N\}$ by I. The goal of this part of the model is to calculate the orientation statistics within a subregion of the image. The resolution of analysis, ρ_σ, at spatial scale σ, determines the sampling rate of a sublattice of points $I_\sigma \subset I$ using the output of filters :

$$I_\sigma = \left\{((i-0.5)\rho_\sigma, (j-0.5)\rho) : 1 \leq i \leq \frac{X-1}{\rho_\sigma}, 1 \leq j \leq \frac{Y-1}{\rho_\sigma}\right\}$$

Now let $N(x, y)$ define a circular neighbourhood, radius P_σ, around each point in the sublattice, I_σ:

$$N_\sigma(x, y) = \{(i, j) : (x - i)^2 + (y - j)^2 < P_\sigma^2\}$$

where $(x, y) \in I_\sigma$. This now defines a set of circular neighbourhoods distributed throughout the image on a grid. The two parameters used in this section are ρ_σ, the sampling rate of the sublattice, and P_σ, the radius of the local region of integration. The values of these parameters are constrained firstly by the size of the filters measuring local orientation estimates (σ) and secondly by the requirement of the model to be *scale-invariant.*

The first constraint is that it is meaningless to use a sampling rate or a neighbourhood size which is too small in comparison to the filter size σ. The maximum sampling rate was set to be $\rho_{1.0} = 8$ pixels (i.e. a 32×32 sublattice) at the finest spatial scale ($\sigma = 1.0$ pixel) and set $P_\sigma = \rho$ as a reasonable neighbourhood size.

The second constraint, scale invariance, has been demonstrated psychophysically by Nothdurft (1985b). Subject discrimination of texture pairs is impaired if the spacing of elements is increased but micro-pattern size remains constant. If both are increased *proportionally* segregation is not affected. In effect this requires that the sampling rate be reduced in proportion to the spatial scale of analysis:

$$\rho_\sigma \propto \frac{1}{\sigma}$$

Define ρ at spatial scale σ as ρ_σ. Since $\rho_{1.0} = 32$:

$$\rho_\sigma = \frac{32}{\sigma}$$

Moving on to consider the partitioning of blobs: from the last section the set of the centroids of all zero-bounded regions in image[2], C, is:

$$C = \{\{(\mathrm{cx}_k, \mathrm{cy}_k)\}, 1 \leq k \leq B\}$$

where $(\mathrm{cx}_k, \mathrm{cy}_k)$ is the centroid of the kth blob and B is the total number of blobs in the image. The set of all blobs centroids falling in a neighbourhood around (x, y) is:

$$C_N(x, y) = N(x, y) \cap C$$

[2]The following is for an image at an individual spatial scale but σ subscript is omitted for clarity.

This defines the sets of blobs that that are used to calculate local texture statistics. The following two sections describe how the mean and variance of blob orientations is calculated.

2.4.2 Mean orientation

Given a set of n estimates of orientation $\Theta = \{\{\theta_k\} : 1 \leq k \leq n, 0 \leq \theta \leq 180°\}$, taken from a region of an image, what is the best way to estimate their overall orientation? One could calculate the arithmetic mean using the angle values themselves. The problem is that this method assumes that the values being averaged are measured on a linear dimension so that arithmetic subtraction reflects the difference between values. When the range of orientations exceeds 90° this is not true because orientation is a cyclic dimension. Angles can differ by a maximum of 90°. The mean of a data set ($\bar{\theta}$) minimises the sum of the differences between all members of the data set and itself. For a cyclical dimension such as orientation one requires a function which reflects the difference between two angles.

The set of measures Θ has a true mean orientation $\bar{\theta}$.[3] Each measurement has an error, e, equal to the difference between it and the true global angle. This error is:

$$e_n = \begin{cases} |\theta_n - \bar{\theta}| & \text{if } |\theta_n - \bar{\theta}| < 90° \\ (90° - |\theta_n - \bar{\theta}|) & \text{otherwise} \end{cases}$$

By the principle of least squares, the mean of the data set, $\bar{\theta}$, should minimise the quantity:

$$E_1 = \sum_{k=1}^{n} e_1^2 + e_2^2 + \ldots + e_n^2$$

Because e is discontinuous, an analytic minimisation of E is not possible but it is straightforward to minimise this quantity iteratively, to an arbitrary level of precision, for a particular set of orientation measures. However, a differentiable alternative for e is also considered: the square of the vector product. Assuming all vectors are unit length:

$$\epsilon_n = \sin^2(\theta_n - \bar{\theta})$$

Since the error term is differentiable, as $\sin^2\theta$ is, an analytic solution to the minimisation problem is possible:

[3] A non-uniform distribution of data is assumed, i.e. one where the true value is defined on a cyclical dimension.

$$E = \sum_{k=1}^{n} \sin^2(\theta_k - \bar{\theta})$$

At the minimum value of E, $dE/d\theta = 0$, so:

$$\frac{dE}{d\theta} = \sum_{k=1}^{n} \sin 2(\theta_k - \bar{\theta}) = 0$$

It follows that:

$$\sum_{k=1}^{n} \sin 2\theta_k \cos 2\bar{\theta} = \sum_{k=1}^{n} \cos 2\theta_k \sin 2\bar{\theta}$$

and hence:

$$\bar{\theta} = \frac{1}{2} \tan^{-1} \left[\frac{\sum_{k=1}^{n} \sin 2\theta_k}{\sum_{k=1}^{n} \cos 2\theta_k} \right] \tag{2.3}$$

This expression gives a value of θ which is guaranteed to yield an extremum of the value of E. This angle may, however, maximise or minimise E; in the latter case θ will be 90° greater than the correct mean orientation. Strangely equation 2.3 alone has been used to estimate global orientation (Kass and Witkin, 1985). Particular problems will occur at angles around 45° where $\sum_{i=1}^{N} \cos 2\theta_i \rightarrow 0$ and 0° where $\sum_{i=1}^{N} \sin 2\theta_i \rightarrow 0$. When summing the individual sines or cosines of a distribution with a mean of around 45°, the sign of these summations is likely to flip randomly, which will cause an estimate of the mean angle to flip between θ and $\theta + 90°$. To resolve this ambiguity one must also evaluate the second derivative, $d^2E/d\theta^2$:

$$\frac{d^2E}{d\theta^2} = -2 \sum_{i=1}^{N} \cos 2(\theta_i - \bar{\theta})$$

then for a maximum:

$$-2 \sum_{i=1}^{N} \cos 2(\theta_i - \bar{\theta}) < 0$$

This expression gives an unambiguous criterion for determining whether an angle is the mean orientation.

Finally, returning to the notation from the previous section, and incorporating the measure of the reliability of each orientation estimate ($\mathrm{Rel}(x, y)$), the orientation extremum, $\theta_e(x, y)$, of the blob set in $N(x, y)$, is defined as:

$$\theta_e(x,y) = \frac{1}{2}\tan^{-1}\left[\frac{\sum_{(i,j)\in N(x,y)} \mathrm{Rel}(i,j)\sin 2\theta(i,j)}{\sum_{(i,j)\in N(x,y)} \mathrm{Rel}(i,j)\cos 2\theta(i,j)}\right]$$

and the corresponding mean orientation, $\bar{\theta}(x,y)$, as:

$$\bar{\theta}(x,y) = \begin{cases} \theta_e(x,y) & \text{if } -2\left[\dfrac{\sum_{(i,j)\in N(x,y)} \mathrm{Rel}(i,j)\cos 2(\theta(i,j)-\theta_e(x,y))}{\sum_{(i,j)\in N(x,y)} \mathrm{Rel}(i,j)}\right] < 0 \\ \theta_e(x,y)+90° & \text{otherwise} \end{cases} \tag{2.4}$$

Let the measure made of $\bar{\theta}(x,y)$ at scale σ be denoted $\bar{\theta}_\sigma(x,y)$.

2.4.3 Orientation variance

Given the local mean orientation the variance of texel orientation in a region may be calculated using:

$$\psi(x,y) = \frac{1}{\nu(x,y)-1}\sum_{(i,j)\in N(x,y)} \sin^2(\theta(i,j)-\bar{\theta}(x,y))$$

Where $\nu(x,y)$ is the number of blobs in each subregion: $\nu(x,y) = \mathrm{Card}(\Theta_N(x,y))$.

Finally a measure of the mean orientation variance in an image is:

$$\Psi = \frac{\sum_{(i,j)\in I_\sigma} \psi(i,j)\rho^2}{XY}$$

Where X and Y are the image dimensions,ρ the sampling rate and I_ρ the set of all lattice points (defined in Equation 2.4.1). This measure is calculated at each spatial scale to give Ψ_σ.

2.5 Selection of spatial scale

Given a multi-local statistical description of the orientation statistics in an image at a particular spatial scale, *which spatial scale should be used?* As discussed previously one problem with many models of texture processing is that the task they are to perform is not defined in enough detail to evaluate their performance as a function of spatial scale. However, because the goal of this model is to produce locally smooth orientation fields

Parameter	Symbol	Value
Number of filter scales	S	7
Filter standard deviations	σ	1-8 pixels
(Filter scales)		(40.7-5.0 c.pi.)
Number of filter orientations	T	12
Filter orientations	θ	$0° - 135°$ in $15°$ steps
R^2 smoothing parameter		2σ
Neighbourhood sampling rate	ρ_σ	$32/\sigma$
(Sampling grid sizes)		(32×32 to 4×4)
Neighbourhood size	P_σ	ρ_σ pixels

Table 2.1. Summary of model parameters

and because a measure of field orientation variance Ψ_σ is available, a criterion for the selection of scale can be formulated:

$$\sigma_{opt} = \sigma : \min_{\sigma=0,1,...S}[\Psi_\sigma]$$

where S is the total number of spatial scales. The addition of this criterion means that the operation of the model is entirely automatic.

2.6 Summary

To reiterate, the model operates as follows. A set of texture element descriptions are generated at each spatial scale using the adaptive filtering scheme. At each scale, the set of texels is divided according to the neighbourhood they fall in; local mean orientation and orientation variance statistics are computed for each neighbourhood. Mean orientation variance is minimised to select a single scale of analysis. The final representation consists of a multi-local set of estimates of mean orientation and orientation variance.

Key parameters of the model are defined in Table 2.1.

The model shares components with other models of early visual processing, and texture processing. An assessment of the importance of individual components of the model, follows:

- Filtering. The important aspect of the filtering operation is that it is *oriented*. Details of the Difference-of-Gaussian are unimportant (Gabors would do).

- Filter energy calculation. The squaring component is not essential: any operation which discards the sign of the output would do (e.g. the absolute value). The

importance of the Gaussian smoothing component depends on the task. For contour integration it is essential, but for most other operations it is not.

- Pointwise selection of filter output. Essential. The combination rule must have an inhibitive component across orientation.

- Texel isolation. The use of thresholding (rather than zero-crossings, for example) is important because the texels produced are larger and have measurable mass, location, etc. The actual value of the threshold could be varied from around 0.5 to 2.0 grey level standard deviations, without significantly affecting performance.

- Texel description. The image description scheme must produce estimates of mass, length, width and orientation.

- Sampling of texels. The number of samples taken per image must vary in inverse proportion to the filter s.d. The minimum number is not essential.

- Orientation statistics. Calculation of the mean orientation and orientation variance could be achieved a number of other ways (e.g. minimisation) but must combine a number of measures to calculate it. The peak orientation will not do for the local average orientation.

2.7 Applications to specific visual tasks

In this section the performance of the model is considered on four visual tasks which require the integration of orientation estimates. The unaltered model, using the parameters described, was run on all four tasks to generate a set of multi-local estimates of mean orientation and orientation variance. Note that the same scale selection criterion was applied in all cases: minimal mean local orientation variance.

2.7.1 Image flow

The application of the adaptive filtering model to this task is illustrated using vector fields of the local orientation estimates, examples are of which are shown in Figure 2.6. These are derived directly from the local mean orientation estimates output by the model. Note that the flow fields not only reflect local mean orientation but also preserve structure such as discontinuities in the field, e.g. the region around the human figure in Figure 2.6c.

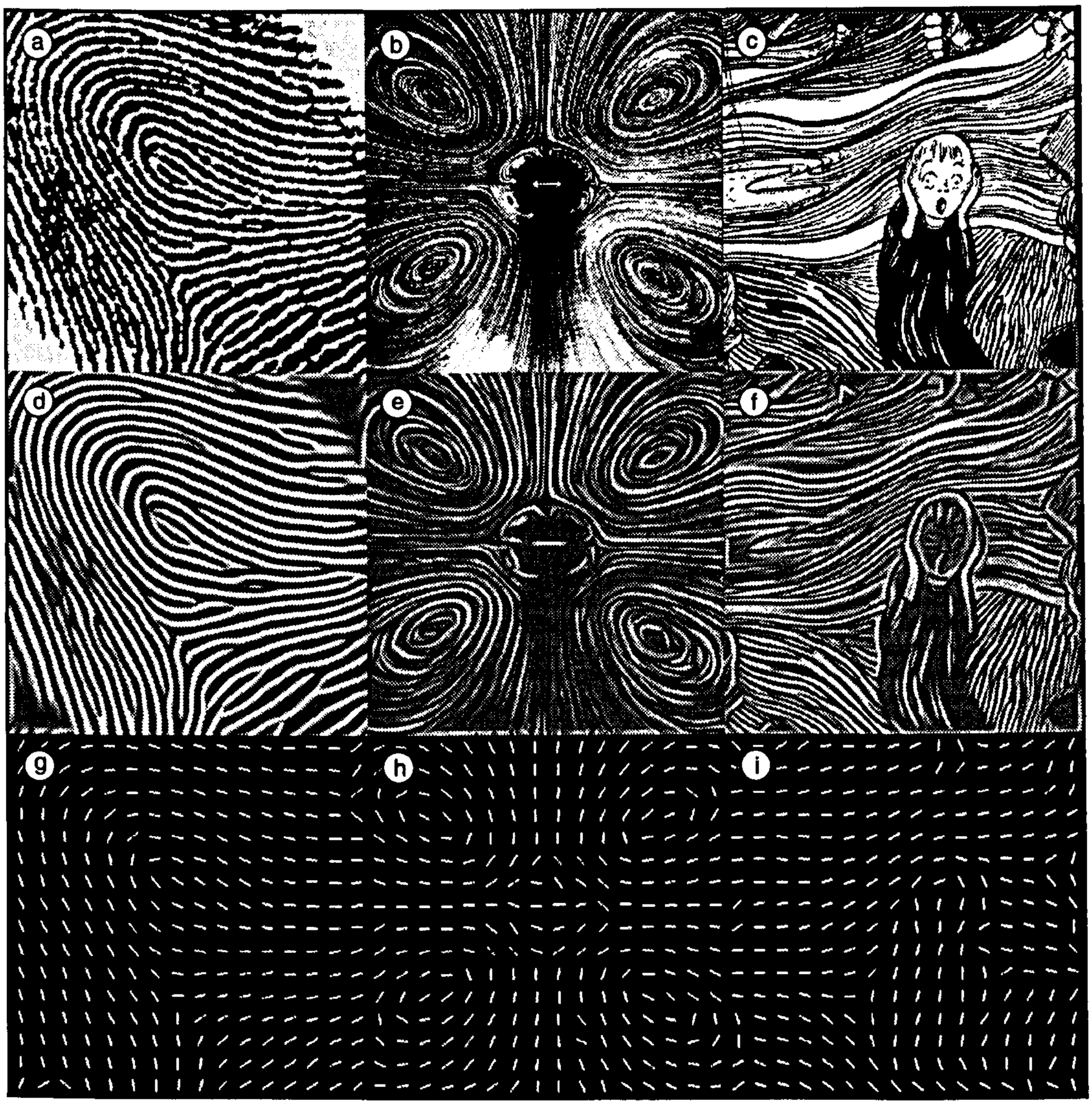


Figure 2.6. Flow fields derived using the model. (a-c) Original images, (d-f) output of filters at optimal scale, and (g-i) flow fields calculated from these images. Subjectively, flow fields correspond well to local mean orientation.

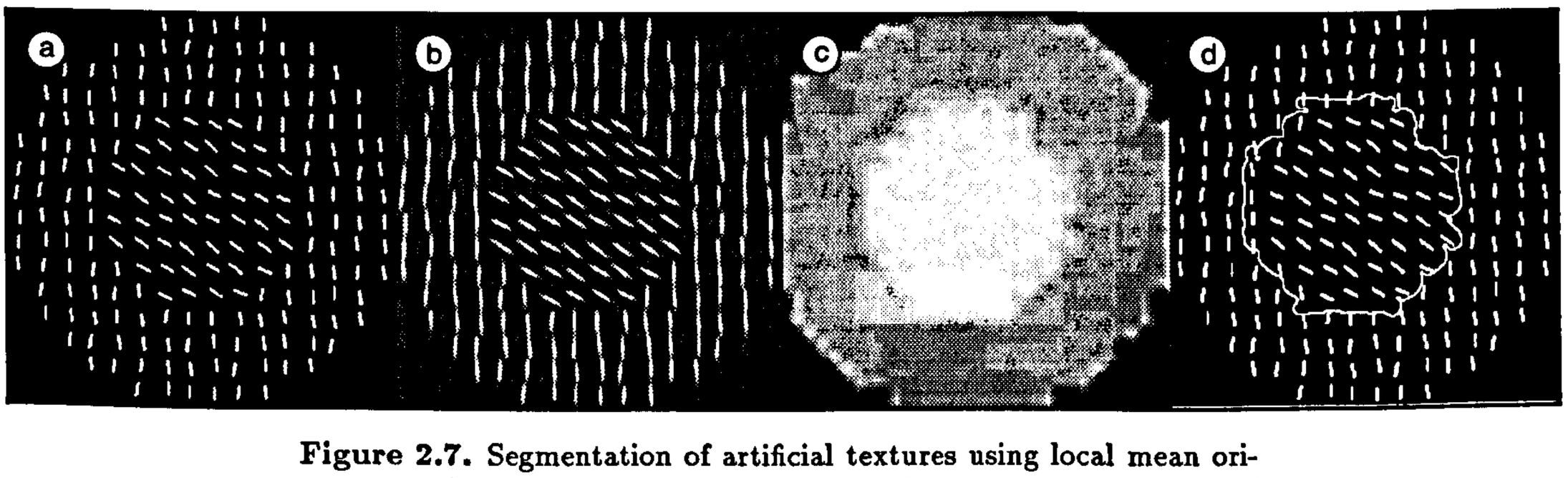


Figure 2.7. Segmentation of artificial textures using local mean orientation. (a) Input image, (b) adaptively filtered image, (c) map of local mean orientation, $P_\sigma(x, y)$, and (d) zero-crossings in the output of (c) superimposed on original image.

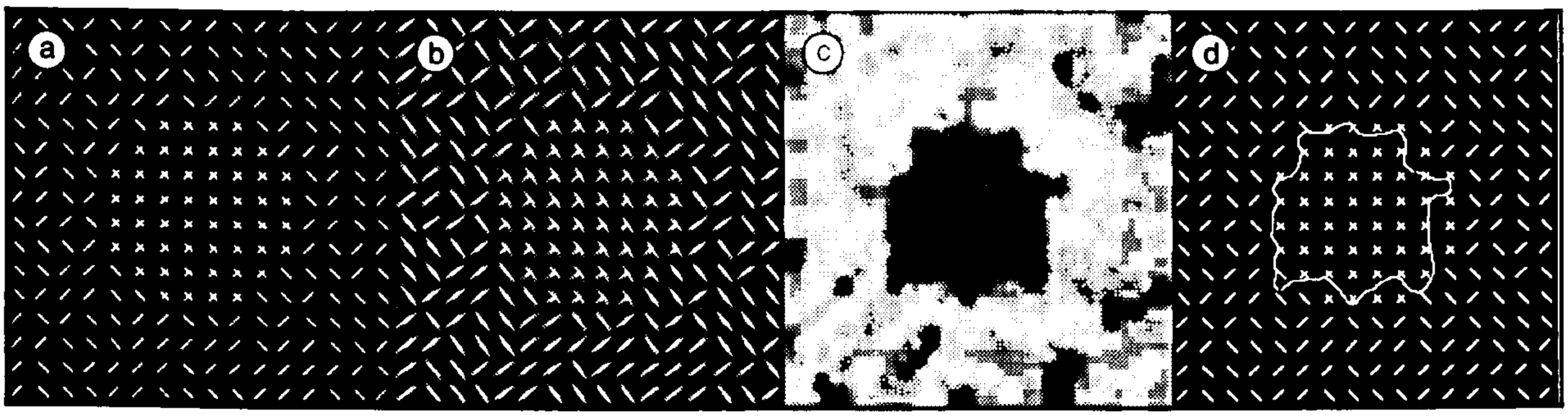


Figure 2.8. Segmentation of artificial textures using local orientation variance. (a) Input image, (b) adaptively filtered image, (c) map of local orientation variance, $\psi_\sigma(x, y)$, and (d) zero-crossings in the output of (c) superimposed on original image.

Figure 2.9. As Figure 2.8 but using local variance to segment embedded natural textures.

2.7.2 Texture segmentation

It is straightforward to extend the basic model described in this chapter to segment scenes according to variation in local orientation statistics. Given the set of mean orientation measures, $\bar{\theta}_\sigma(x, y)$, and orientation variance measures, $\psi_\sigma(x, y)$, we can resize these arrays to the size of the original image (using bicubic spline interpolation). If we wish to highlight spatially distinct regions in these representations image we simply convolve each with a relatively coarse scale Gaussian filter and indicate the zero-crossings. In addition a higher sampling rate was used (4ρ) to ensure high resolution of boundaries. Note that this model overcomes two of the three criticisms outlined above: there are no parameters set by hand (scale is selected automatically), and this operation operates on a representation which is used in a variety of visual tasks. It is beyond the scope of this chapter to present a thorough psychophysical evaluation of this model so simple subjective demonstrations are shown in Figures 2.7, 2.8 and 2.9. The model, given the three segmentation images as input, derives regions that approximately reflect perceived boundaries.

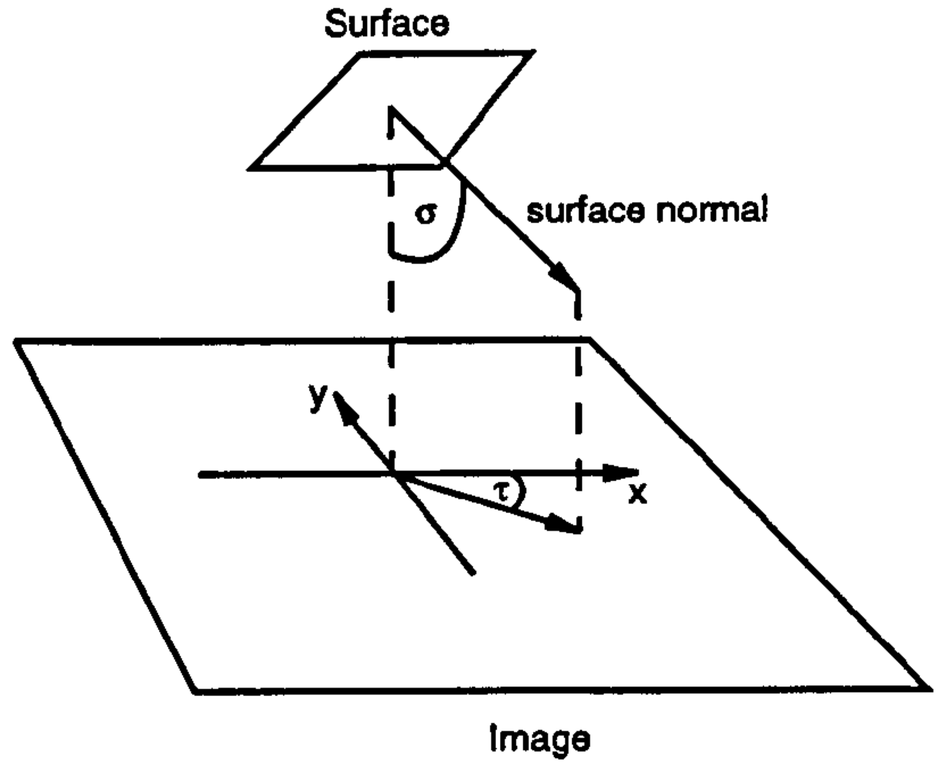


Figure 2.10. Slant and tilt

2.7.3 Shape from texture

Shape is generally defined in terms of surface slant and tilt (Figure 2.10). Slant (σ) is the angle between a normal to the surface and the line of sight. Tilt (τ) is the angle between the projection of the surface normal into the image plane and some arbitrary dimensions in the image plane (usually the x-axis). Tilt defines the direction of slant.

Witkin (1981) uses zero-crossings in the output of Laplacian-of-Gaussian filtered, texture images to derive oriented texels. Given that a set of local orientation estimates are available from these texels, he derives a probability density function (p.d.f.) for image tangent direction under projection, α^*, on a surface with slant, σ, and tilt, τ, as:

$$p.d.f.(\alpha^*|\sigma,\tau) = \frac{1}{\pi}\frac{\cos\sigma}{\cos^2(\alpha^*-\tau)+\sin^2(\alpha^*-\tau)\cos^2\sigma} \tag{2.5}$$

Hence, given a set of tangents grouped into n orientation bands - $A^* = \{\alpha_1^* \ldots \alpha_n^*\}$, the relative likelihood function for σ and τ is calculated as:

$$L(\sigma,\tau|A^*) = \exp\sum_{i=1}^{n}\frac{\pi^{-2}\sin\sigma\cos\sigma}{\cos^2(\alpha^*-\tau)+\sin^2(\alpha^*-\tau)\cos^2\sigma}$$

This expression may be applied to multiple estimates of A^* taken from regions of the image, to derive *local* shape information.

Probability density functions, derived using Equation 2.5, are plotted in Figure 2.11 for different values of σ and τ. The formula expresses two dependencies of texel orientation on surface orientation: texels align perpendicular to the orientation of the surface, and the degree of this alignment is proportional to the degree of surface slant away from the observer. Thus, notice that (a) slant affects the sharpness of the peak in the p.d.f., and

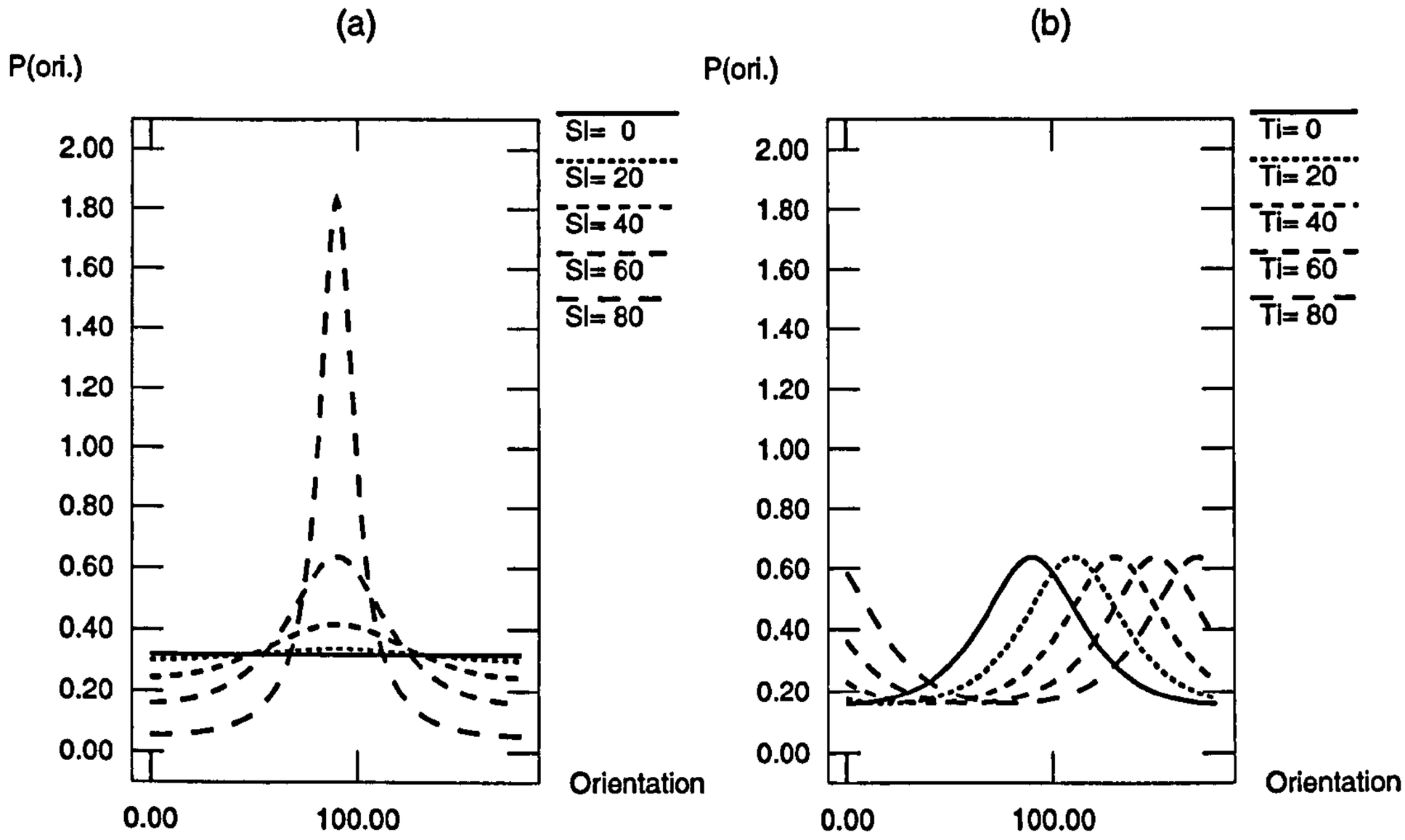


Figure 2.11. Predicted distributions of texel orientations, from Equation 2.5. (a) Shows distributions for a 90° tilted surfaces at various slants, and (b) for a 60° slanted surface at various tilts. Notice that as slant increases, in (a), the width of the distribution increases, and that as tilt increases, in (b), the position of the distribution shifts.

(b) tilt only changes the position of that peak. This is the basis of the use by Blake and Marinos (1990) and Kanatani (1984) of the second statistical moments of texel orientation distributions as an indicator of local surface orientation. Kanatani (1984) used the *Buffon transform*, a measure of the average number of intersections of a straight line dropped on to the texture, which operates on the coordinate vectors of individual texels. The Blake and Marinos (1990) system operates on line elements and is not only more reliable at extreme surface slants but also produces a measure of *reliability* in the surface orientation estimate.

Witkin (1981) uses a maximum likelihood operator to estimate the parameters of Equation 2.5. This involves the solution of a number of nonlinear equations, and it has been suggested that the procedures offered by Witkin (1981) are computationally inefficient (Gårding, 1993).

An alternative method is to extract only the orientation mean and variance which best approximate the given functions, since there is a unique mapping from any combination of mean and variance to the correct p.d.f. The mean and variance statistics are directly available from the adaptive filtering model. The local mean orientation translates directly

to the local tilt:

$$\tau(x, y) = \bar{\theta}(x, y)$$

To estimate slant from local orientation variance, a polynomial approximation can be made to the mapping from local orientation variance to surface slant[4]. This is illustrated in Figure 2.12.

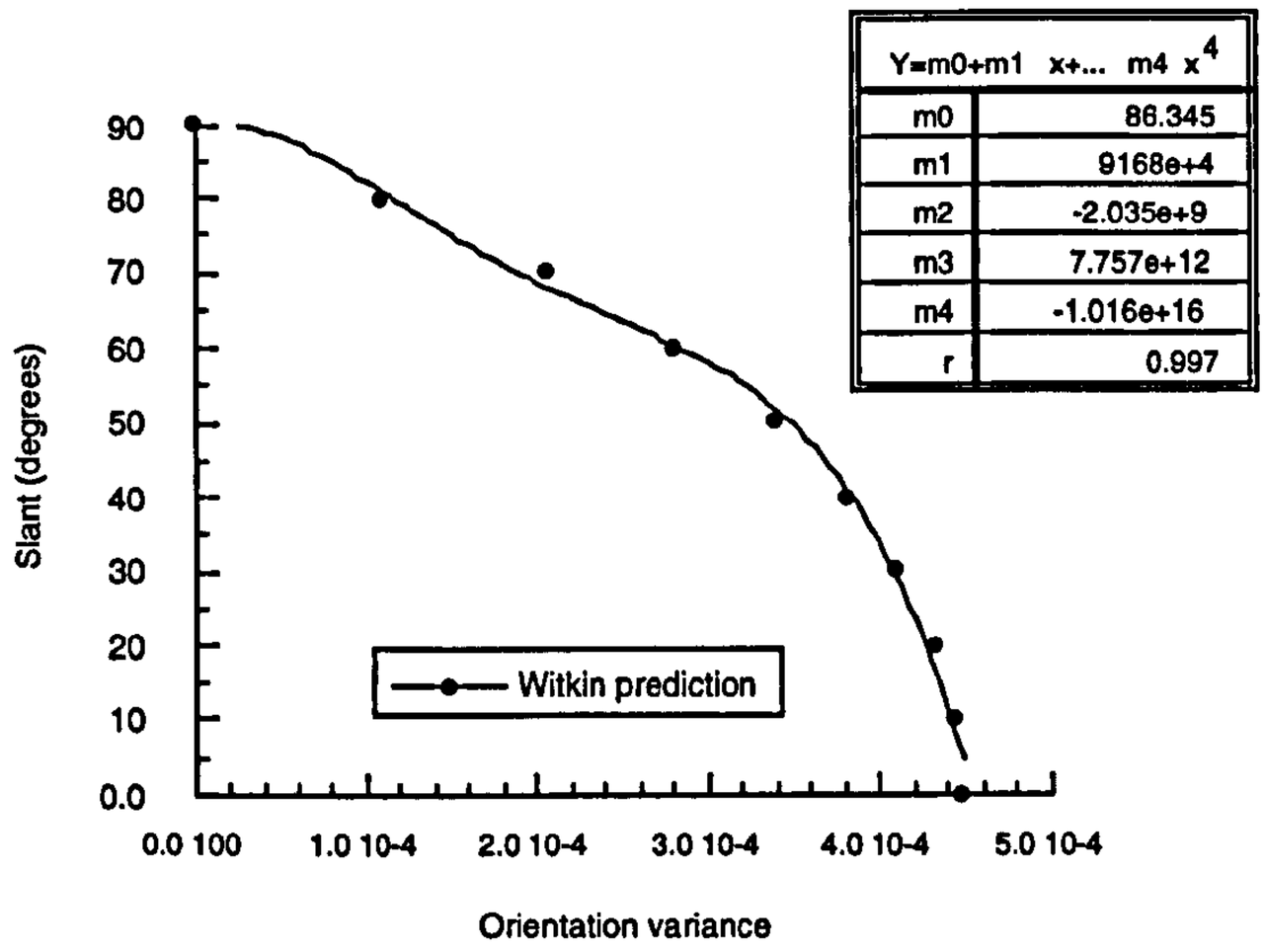


Figure 2.12. The inverse mapping from local variance to slant. Data were obtained by calculating the variance of orientation distributions at different slants, according to Equation 2.5. The solid line is a polynomial fit to the data and account for 99.7% of variance.

To select the scale of analysis the minimum local variance criterion is used. Note that this will automatically select the texel set which maximises local orientation field smoothness. If texture sets tend towards isotropy, as Witkin (1981) assumes, then this procedure will select the most "interesting" scale in terms of deviation from isotropy.

Figure 2.13 shows the operation of the model on two images of Scottish mountains. The derived surface orientation maps are noisy but do approximately reflect local shape.

2.7.4 Contour integration

The perception of image flow in Glass patterns (Glass, 1969) and other vector field textures demonstrates that the human visual system is capable of forming extended curved contours on the basis of a set of local orientation estimates. This principle, termed the law of "good continuation" by the Gestaltists (e.g. Koffka, 1935), has been the subject

[4]A more formal treatment of the relation of Witkin's maximum likelihood estimator to a method of moments scheme, has recently appeared (Gårding, 1993).

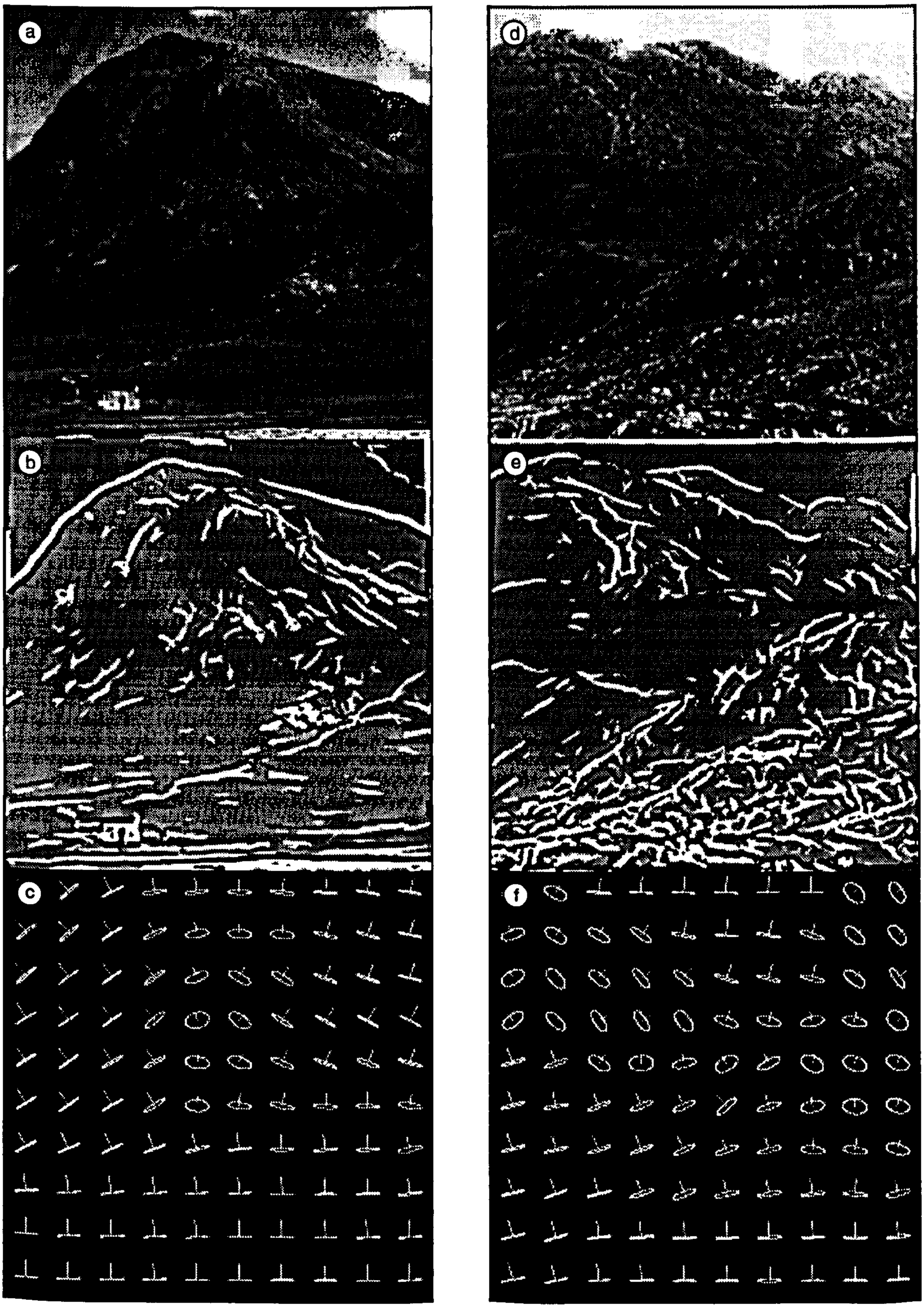


Figure 2.13. Images of (a) Glen Coe and (d) Ben Venue. Local orientation statistics, derived from adaptively filtered versions (b,e), are used to construct a map of local surface orientation, (c,f).

of several psychophysical and computational studies. Straight lines composed of dots are visible when embedded in unstructured fields of dots, depending on the dot spacing and collinearity (e.g Uttal, 1975). Many psychophysical studies of contour detection in noise have been poorly controlled and/or have not examined more than one stimulus dimension. Recently however Field *et al.* (1993) presented psychophysical data from an unusually systematic and controlled study of how the parameters of continuous contours effect their detection within texture fields. Field *et al.* (1993) examined the detection of continuous contours in patterns composed of grids of Gabor patches. Embedded within a field of 244 randomly oriented elements was a set of 12 patches lying on a "jagged" path (e.g. Figure 2.14b). Using a forced-choice procedure the experiment showed that subjects could detect fields containing a path even when the spacing of path elements exceeded element size or when the relative orientation of path elements was as much as $\pm 60°$. The alignment of the elements along the path is critical: orienting elements orthogonal to path direction and allowing variation in patch orientation around the path direction both reduce detection performance. Finally it was found that the relative phase of the Gabors had no affect on path detection.

A number of models have been proposed for the integration of local measures of orientation into coherent contours. They are virtually all based on an operation on the output of a set of oriented feature detectors which favours orientation measures which are "consistent" with each other in a local neighbourhood, i.e. if both orientations are consistent with the hypothesis that a single contour caused them. Models have used "association fields" (Field *et al.*, 1993), "co-circularity support" (e.g. Parent and Zucker, 1989) or "hyper-filters" (Sagi, 1990). Parent and Zucker (1989) measure local orientation with Difference-of-Gaussian filters and then use neighbouring orientations to estimate how consistent the orientation measure is with being a tangent to a curve. This co-circularity support is implemented in a relaxation network. As Field *et al.* (1993) point out, the problem with such iterative "linking" processes, as models of human vision, is that they are extremely time consuming. They also note that many computational systems are not well suited to dealing with contours composed of discrete elements such as Gabor patches or dots. The system that they propose, using "association fields", is illustrated in Figure 2.15. Excitatory links extend between orientation paths that are consistent with each orientation estimate being a tangent to a contour at that point. Inconsistent orientations are inhibited. The major problem with this approach is that the resultant

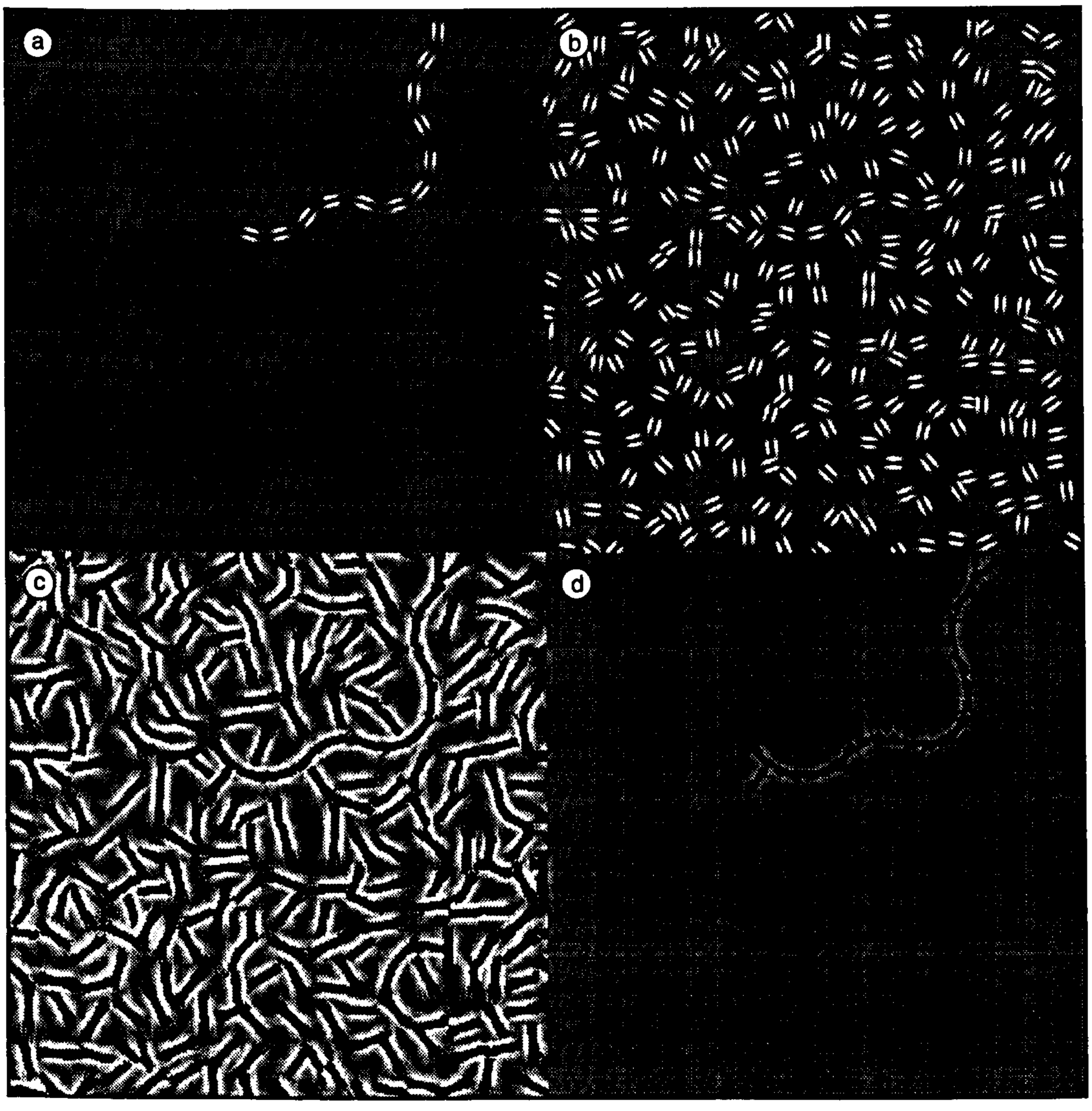


Figure 2.14. Explanation of the path detection task described in Field *et al.* (1993) using the adaptive filtering model. (a) A typical path, with successive elements differing in orientation by $\pm 45°$, that subjects detected when embedded in a field of randomly oriented elements (shown in (b)). (c) Optimal filtered version of (b) derived using the model - note merging of elements in the path. Selection of the largest feature in this image, shown in (d), yields the correct path.

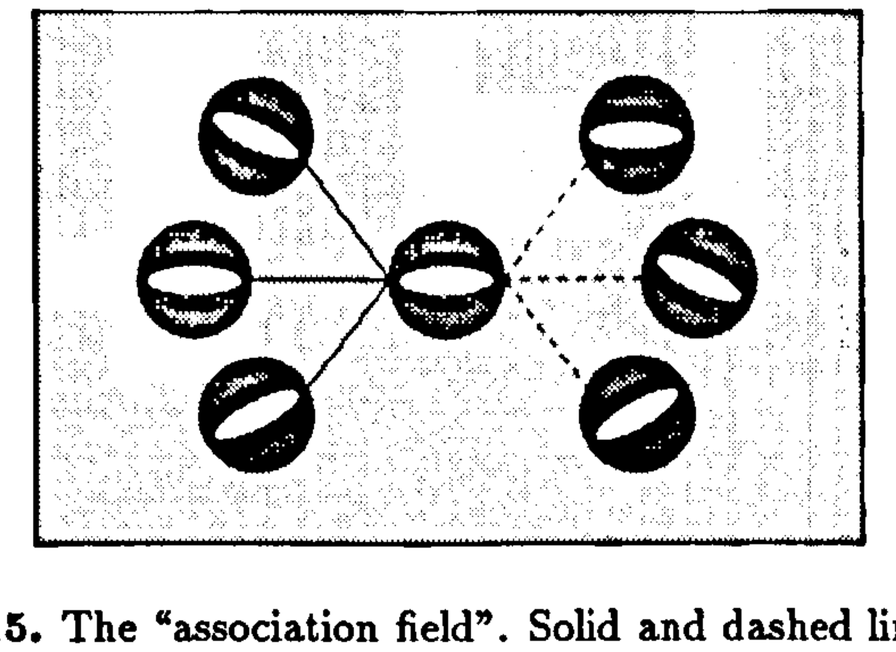


Figure 2.15. The "association field". Solid and dashed lines indicate excitatory and inhibitory connections respectively.

wiring scheme is tremendously complex, given that all that is known from neurophysiology with any confidence is that there are long range excitatory connections between cells with similar orientation specificity (Gilbert and Wiesel, 1989; Grinvald *et al.*, 1989).

In the adaptive filtering model described earlier, it was proposed that filter selection occurs on a point-by-point basis according to the power of a particular oriented filter at that point. Additionally a degree of spatial smoothing is applied to the power using Gaussian blurring. This Gaussian smoothing is consistent with the presence of excitatory interactions between cells with similar orientation selectivity (Gilbert and Wiesel, 1989; Grinvald *et al.*, 1989).

Figure 2.14b shows a typical stimulus from Field *et al.* (1993). Successive elements in the path, shown in isolation in Figure 2.14a, differ by ±45°. Field *et al.* (1993) found that subjects' performance breaks down at around element differences of ±60°. The output of the adaptive filtering model is shown in Figure 2.14c and a large blob corresponding to the path is visible. If one simply selects from this image the feature with greatest mass (using the Watt (1991b) image analysis scheme) the path is isolated, as illustrated in the final section of the figure. This suggests that for stimuli such as these there may be no need to resort to an explanation of performance based on the interpretation of the *output* of spatial filters. Instead the local selection of filter outputs based on the magnitude of their output produces effective aggregation along contours.

2.8 Conclusions

A model has been described for the multi-local representation of orientation information derived from visual texture.

The use of local orientation mean and variance statistics was shown to be useful not only as a way of representing "image flow", but also as a simple method of texture segmentation and deriving local shape-from-texture. The use of adaptive filtering has also been shown to have unexpected advantages for the detection of visual contours. The use of smoothed local energy to select filter response at a point means that some degree of aggregation of filter responses occurs in the direction of the filter orientation.

The following chapters investigate, psychophysically, how similar the model's representation is to the human visual representation of orientation. The model throws up a number of unresolved questions, which will be considered in the course of this thesis:

- Can we integrate a set of discrete measurements of orientation at all? If so, what

evidence is there that the human visual system uses a measure of mean orientation rather than, for example, the most frequently occurring orientation? (Chapter 3).

- How do we know that the filter scale-selection problem is a real one? Does filter selection really affect the accuracy of orientation estimates? Is oriented filtering necessary for making local estimates of orientation? What evidence is there that humans outperform the accuracy of isotropic mechanisms? (Chapters 4-5).

- Does the neighbourhood scheme proposed in this chapter allow for the representation of interesting global orientation structure? Will it retain potentially interesting anomalies in the field? (Chapter 6).

- Mechanisms based on spatial filtering are sensitive to certain contrast effects: how does the contrast polarity of the oriented elements making up a texture affect human processing compared to how it affects the model (Chapter 7).

3 The representation of average orientation

3.1 Overview

This chapter examines how observers estimate the statistics of oriented textures. Specifically it tests the hypothesis that, given a set of separate measures of orientation, the visual system uses the arithmetic mean of those measures to represent the central tendency of the set.

Experiment 3.1 attempted to show that observers can combine multiple orientation estimates when asked to judge the average orientation of a texture composed of discrete elements. Subjects were presented with patterns composed of Difference-of-Gaussian patches, whose orientations varied around some mean value, and asked to make a clockwise-anticlockwise judgement as a function of the orientation variability of the pattern. The level of performance achieved indicated that subjects must be integrating a number of separate measures together to perform this task.

Given that some form of averaging is taking place, Experiments 3.2-3.5 used *asymmetrical* distributions of orientations to try to separate the predictions from different models for estimating average orientation. Stimuli were composed of two spatially intermingled sets of DoG patches, each set having Gaussian distributed texture element orientations. The threshold separation of the mean of the two sets was then determined for a variety of tasks.

Discrimination of these textures from others composed of sets with *equal* mean orientations was well predicted by global variance discrimination. The judgement of which set contained more texture elements elevated thresholds and suggested (a) that the skew of orientation distributions is unavailable and (b) that the resolution limit of the two distributions is a minimum separation of texel sets for independent representation of their properties.

Threshold offsets for judgement of average orientation were expressed as shifts of

four candidate features for coding the central tendency of texel orientations. They were then compared to similar thresholds for single texel sets. Results indicated that average orientation is assigned to the centroid of a set of orientation measures.

However a system which simply averages over the whole orientation range to estimate average orientation cannot represent multiple sets. Experiment 3.6 investigated whether observers are capable of averaging over a limited *range* of orientations which would possibly indicate that some form of filtering of orientation distributions occurs to allow for background "noise".

3.2 INTRODUCTION

Most psychophysical studies of texture perception, largely influenced by the work of Julesz (e.g. Julesz, 1981), have concentrated on the extraction of texture *boundaries*. This work presupposes that texture is primarily useful as a supplement to contrast information for the detection of surface discontinuities. Experimentally, boundaries are usually defined locally as differences between attributes of adjacent texel pairs. This approach assumes that, in general, textural cues will be available from local differences between texture elements (texels). However, attribute information derived from natural images is invariably noisy and must require spatial combination to be useful. In other words, approaches to texture processing that assume that simple local differences between texels suffice are probably misguided, since averaging of texture attribute information is required for dealing with natural images.

If such integration does occur, what useful information might be derived from it? This, of course, depends on what is to be done with it. However, the extraction of global texture attributes (particularly orientation) is known to be useful in establishing a number of different surface properties. Kass and Witkin (1985) used such statistics to estimate the formative processes a texture had undergone. Similarly estimates of local surface shape require integration and calculation of (usually second) moments of local orientation statistics (e.g. Witkin, 1981; Blake and Marinos, 1990). Additionally, segmentation might be achieved not only through local differences in attributes but also in differences between integrated attribute statistics (e.g. difference in mean orientation: Voorhees and Poggio, 1987). Little is known, however, of human perceptual performance where textural moments are explicitly varied or of what coding strategies are used to describe texture statistics.

Since orientation appears to be so useful for deriving surface properties and is one of the proposed textons for which segmentation performance is independent of local luminance fluctuation (Nothdurft, 1990), it is concentrated upon as the texel attribute of interest.

Marr (1982) observed that, in order to understand a complex process such as texture perception, the formation of intermediate representations is required. What are the primitives for representing integrated orientation statistics? Recent observations suggest an analogy between the visual processing of luminance features and textural boundaries. It has been shown that texture segregation is sensitive to the rate of change of texel orientation across space (the *structure gradient*) at field boundaries (Nothdurft, 1985a;

Nothdurft, 1985b; Nothdurft, 1991b). Landy and Bergen (1991) used spatially band-pass textures to show that this effect is independent of the local density of texels. In addition, Sagi (1990) found a non-monotonic relationship between performance on a visual search for a vertical target and the number of (horizontal) distractors. Data were well explained using "hyper-filters", which integrate local orientation measures over a restricted area. Input is independent of spatial frequency, and filters indicate significant changes of these measures in space (Sagi, 1990; Fogel and Sagi, 1989).

These studies suggest that performance on visual search and texture segregation tasks, using orientation cues, may be well explained by mechanisms which extract differences between spatially integrated local orientation measures. Together they strongly implicate a further stage of combination and processing of orientation information beyond the accepted representation at V1 (Hubel and Wiesel, 1967).

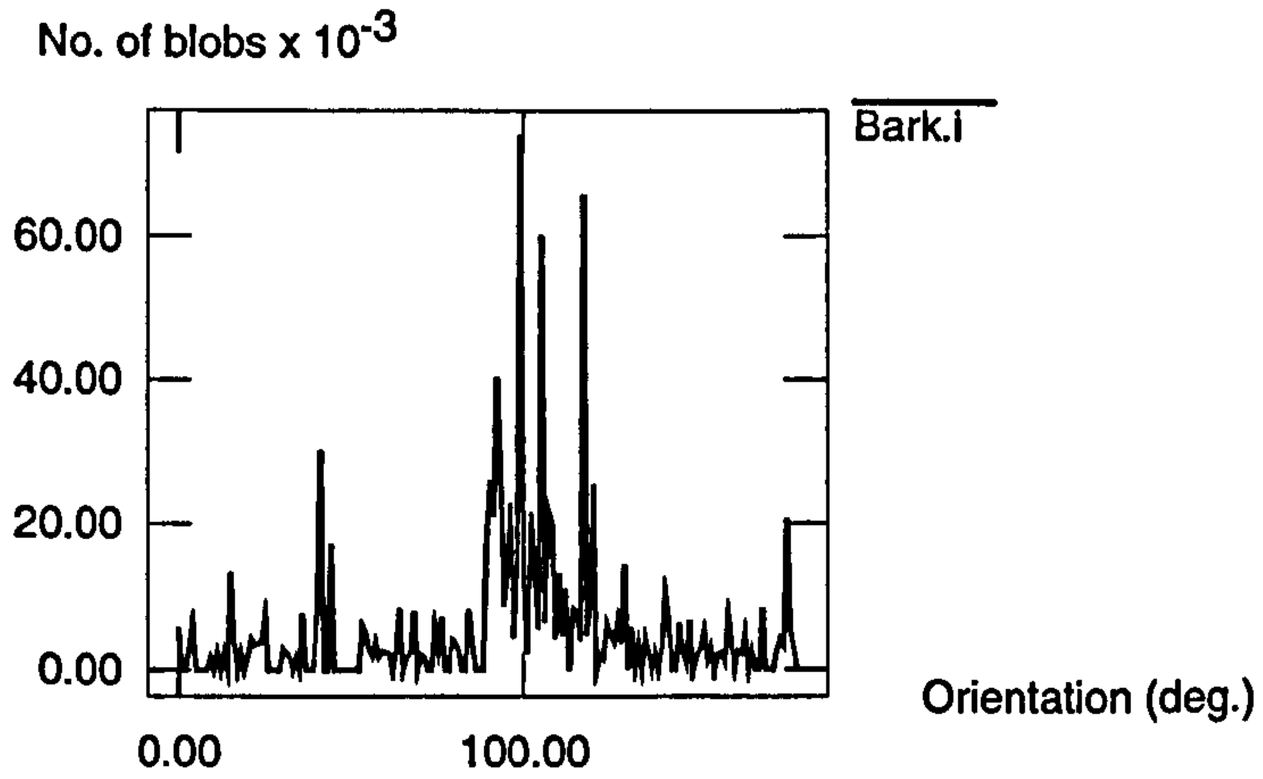


Figure 3.1. (a) Image of tree bark, a strongly oriented texture, (b) texels isolated from the $\nabla^2 G$ filtered version of (a), (c) orientation histogram of blobs from (b). Which statistics of this distribution are used in texture processing?

Consider Figure 3.1 which shows (a) a highly oriented texture, (b) a set of candidate

texture elements (derived using the thresholded output of $\nabla^2 G$ filters) and (c) the orientation histogram of these blobs. One is capable of correctly estimating that the average orientation of the original texture is approximately vertical. However, in the case of this highly oriented texture, there is a great deal of noise present in the orientation histogram. If one were interested in an orientation representative of the overall direction of surface creases (for example, to position the cutting edge of a tool on the surface) one would have to use some form of *combination* of measures to calculate it. Individual texels/blobs are too unreliable. Similarly, to use texel orientation for surface *slant* estimation requires that one evaluates the deviation of the orientation distribution from isotropy, usually using the second moment of the orientation distribution (Blake and Marinos, 1990). Again this requires integration of a set of orientation measures. The question that this chapter addresses is this: how does the human visual system calculate such statistics, particularly the *central tendency* of texel orientation distributions?

Observation of how a system's performance deteriorates with the addition of noise is a powerful approach for understanding that system (see, e.g. Barlow, 1980). By employing textures composed of separable texels one may examine the effect of noise explicitly added to the orientation of components. This approach to understanding the perception of average orientation of textures has already been applied elsewhere (Watt, 1991b). When the orientation of texels varies around some mean value, judgement of average orientation is limited only by the variance of texel orientations and by some constant internal error of representation. By further supposing that orientations are averaged, an estimate of the number of texels employed can be made. Such a model accounts for the data presented and suggests that the mean orientation of a number of orientation measures was available to subjects.

To reiterate, Watt (1991b) offers preliminary evidence that integration of texel orientation occurs in the judgement of average orientation of a texture field. Furthermore this work suggests that the arithmetic mean of texel orientations predicts thresholds well. Experiment 3.1 uses a similar experimental design but uses band-pass textures to confirm that subjects average across multiple orientation measures to estimate average orientation.

The remainder of the chapter explicitly compares alternative strategies for coding average orientation. Is averaging of orientations required at all, or could an alternative feature (such as the most numerous orientation) suffice? In addition, the representation of multiple oriented sets is investigated. The understanding of these issues will give insight

into the representation and primitives used, and how they are calculated.

In the experiments reported in this chapter all stimuli were spatially unstructured single patches of oriented texture. The experimental procedures and models employed explore an analogy between the processing of luminance features and texture. Assume that some estimate of the generative orientation probability function is formed through averaging of local texel attributes, i.e. an orientation histogram is constructed. It is hypothesised that the extraction of statistical moments (e.g. central tendency) of this representation will occur in a similar way to the extraction of moments from retinal luminance distributions. In the case of the central tendency of a retinal distribution two stages are involved in most models: the integration of luminance values through a blurring function and the accentuation of some moment of the resulting distribution (e.g. the second derivative). It is proposed that in forming an estimate of the central tendency of an orientation distribution, similar processes may be used.

The analogy between the processing of orientation and luminance distributions has been explored elsewhere. Keeble and Morgan (1993) had subjects discriminate between sinusoidally modulated, and uniform orientation p.d.f.s, as a function of the amplitude and frequency of the sinusoid. Subjects perform optimally with distributions modulated at one cycle per 180°, which suggests the task is performed by convolving the p.d.f with an "orientational filter". This filter had an excitatory zone 55° wide, flanked by inhibitory zones. Just as this procedure is strongly analogous to measurement of contrast sensitivity using luminance gratings, the experiments in this chapter, examining the coding of average orientation, are strongly analogous to experiments performed to examine *visual location* in the spatio-luminance domain.

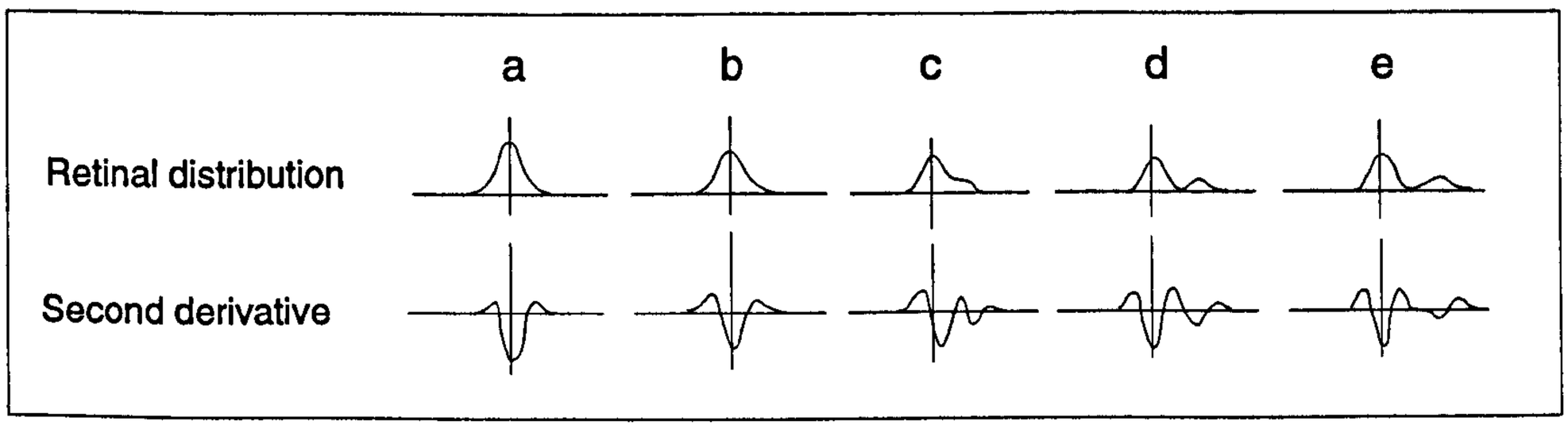


Figure 3.2. The upper row shows the retinal light distributions of typical stimuli from Watt and Morgan (1983a). The lower row shows the second derivatives of the distributions from the upper row. (a) Shows the distributions from two superimposed bright bars with luminance rations of 2:1. (b-e) The effect of increasing the separation of the bars.

Watt and Morgan (1983a) employed tasks using spatially combined bright bars to compare the predictive power of four visual location models. By varying the relative brightness of component bars, the asymmetry of the compound feature presented could be varied. Using these stimuli the threshold offset of the mean positions of the bars was determined for vernier acuity and resolution tasks. Figure 3.2 shows the retinal light distributions, and their second derivatives, of a typical stimulus from the experiment. The distributions from two superimposed bars, with luminance ratios of 2:1, are shown in Figure 3.2a. Figure 3.2b-e shows the effect of increasing the separation of the two bars. At the resolution point (Figure 3.2c) the number of zero-crossings in the second derivative increases from two to four. A model based on the extraction of zero-crossings in the second derivative best accounted for the subjects' accuracy at estimating the position of these asymmetrical luminance distributions. A centroid model also performed well and has been implicated elsewhere in deriving the location of dot clouds (Whitaker and Walker, 1988) and flanked bars (Badcock and Westheimer, 1985). Peak and threshold edge models produced poor fits to the data.

The Watt and Morgan (1983a) paper was used as a framework for experimental method and the models they examined were adapted to process texel distributions. In particular the extraction of the central tendency of orientation distributions and the resolution of compound distributions were examined. Rather than using bars defined by a distribution of luminance in space, the orientation of texels was determined by two probabilistic distributions. Textures will have *orientation distributions* similar to the upper row of Figure 3.2. By shifting the relative mean orientations of the two distributions an asymmetry can be constructed, analogously to the spatio-luminance case.

The textures that were used consisted of two intermingled sets of band-pass oriented texels (see Figure 3.6 for examples). The orientations of elements were randomly drawn from Gaussian distributions ($\sigma = 6°$) and the relative number of texels in each set was systematically varied. Using an adaptive psychophysical procedure the threshold offset of the mean orientations of the two sets was determined for three tasks. They were:

- A "resolution" judgement: subjects identified the patch that was composed of sets with different mean orientations.
- A judgement of which set within one texture patch was more numerous.
- A comparison of the average orientation of the patch with vertical.

Several models were compared to determine the cues used to make the judgements. Performance for judging the average orientation of asymmetrical distributions (Experiments 3.4-3.5) was used to separate predictions from the models, as was the case in Watt and Morgan (1983a).

3.3 Methods

3.3.1 Subjects

The author and five observers naive to the purposes of the experiments served as subjects. Three (SCD, PMC and DFS) were experienced psychophysical observers, while the others had limited experience. DFS and SCD were corrected-to-normal myopics and DFS also had a slight (less than 0.5 D) corrected astigmatism.

3.3.2 Apparatus

All stimuli were presented on a Formac ProNitron 80.21 colour monitor with a frame refresh rate of 80 Hz. The screen was viewed binocularly with natural pupils at a distance of 49 cm, and had a mean luminance of 92 cd/m^2. Subjects fixated the centre of the display, indicated using a pre-stimulus fixation marker.

3.3.3 Stimuli

In order to restrict orientation information to an individual spatial scale, textures with relatively narrow-band spatial frequency characteristics were used. Textures were composed of small patches with luminance modulated by a Difference-of-Gaussian function. Many of these DoG elements were randomly distributed throughout the image and added together. They appeared on a mid-grey background. The DoGs making up the textures had a luminance profile given by:

$$L(x,y) = L_0[1 + W(x,y)] \tag{3.1}$$

where:

$$W(x_t,y_t) = A\left[e^{-x_t^2/2\sigma^2} - (\frac{1}{2.23})e^{-x_t^2/2(2.23)\sigma^2}\right] e^{-y_t^2/2(3.2\sigma^2)} \tag{3.2}$$

$\sigma = 3.48$ min. arc and A is a scaling constant. x_t and y_t are the translated and rotated coordinates:

$$x_t = (x - \mu_x)\cos\phi + (y - \mu_y)\sin\phi \tag{3.3}$$

$$y_t = (y - \mu_y)\cos\phi - (x - \mu_x)\sin\phi \tag{3.4}$$

μ_x, μ_y represent the translation and ϕ the rotation of the function (this is an adapted sensitivity profile given by Wilson and co-workers (Wilson and Gelb, 1984; Phillips and Wilson, 1983) for the 4 cycles per degree channel.) The DoGs were clipped to ±3.0 standard deviations from their centres. Overlapping patches were numerically added. The whole image was stored with floating point precision before being normalised to a range appropriate to the display hardware for presentation. Differing degrees of overlap between texels effectively randomised the contrast range from trial to trial.

The texture patch contained a Gaussian random spatial distribution of texels ($\sigma = 50$ pixels), centred on the middle of the display. The texture fitted into a central window of 256 pixels square (10.0 degrees square).

3.3.4 Procedure

A Macintosh IIfx microcomputer generated and presented the stimuli, and recorded subjects' responses. The subjects' task was always a binary decision involving one (Experiment 3.1 and 3.3-3.6) or two (Experiment 3.2) intervals. Unless stated otherwise, textures were presented for 100 ms. For the two interval experiments, the first texture was presented in the centre of the display, followed by a 750 ms. delay followed by the second texture. In the single interval experiments the texture was presented, followed by a 750 ms. delay. The instructions given to subjects in the single interval experiments, which all involved the judgement of average orientation, were to try and judge the "overall", or most representative, orientation of the patch. Subjects indicated their decisions by depressing one of two keys on the computer keyboard. APE, an adaptive method of constant stimuli (Watt and Andrews, 1981), was used to sample a representative range of relative numerosities of the two sets[1]. Probit analysis was applied to the response data to estimate the standard deviation of the psychometric function. Three runs of 64 trials each were undertaken for all conditions, unless stated otherwise. The data points presented are the arithmetic mean of thresholds from the three runs and error bars are an estimate of the standard error of these data.

[1]The version of APE used differed slightly from the procedure described in this reference. The fitting of the response distribution with the cumulative normal error model was performed *every* trial, and not after each *block* of trials, as described in the Watt and Andrews (1981) paper. This addition only serves to increase the efficiency of the algorithm.

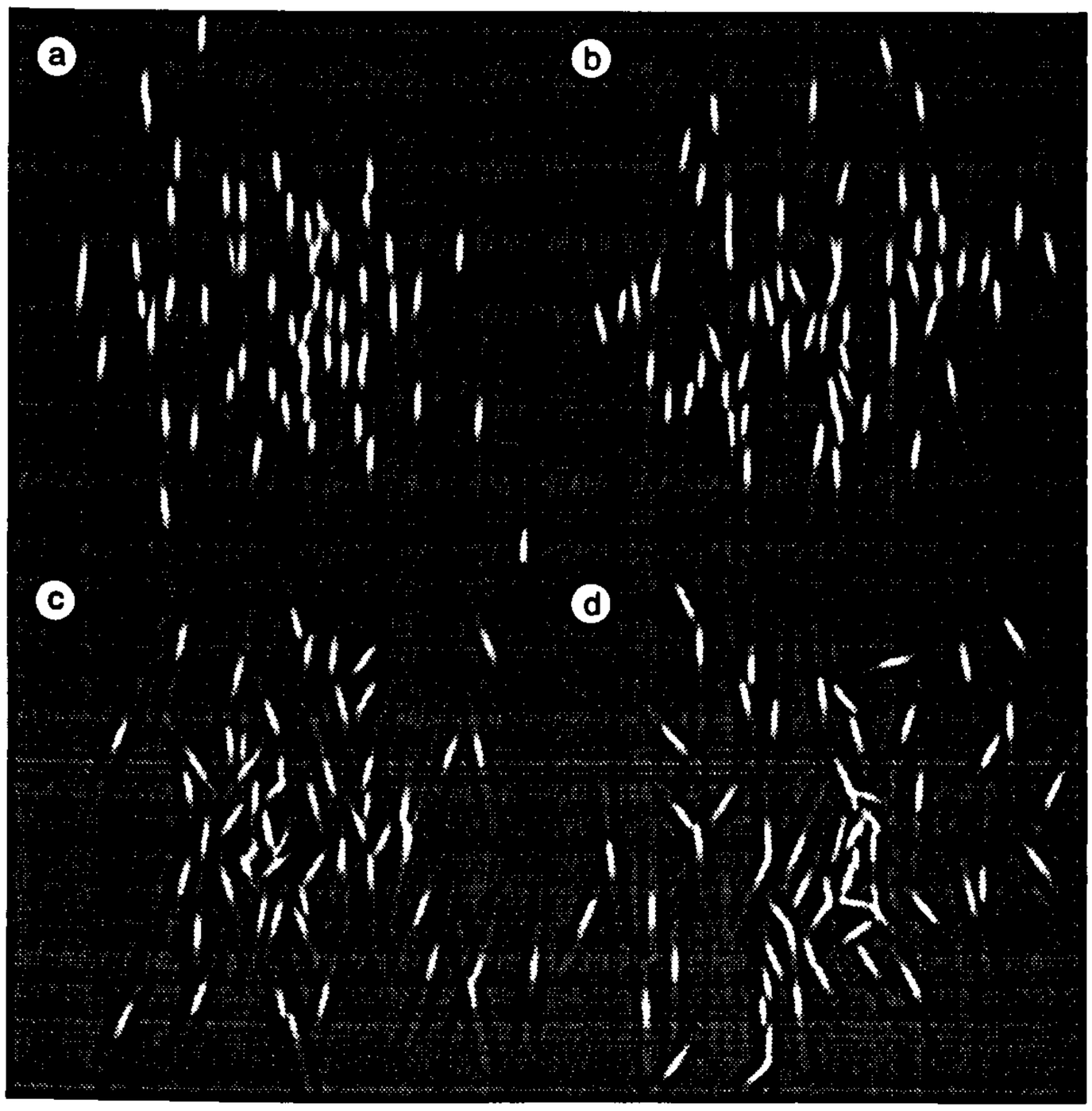


Figure 3.3. Examples of the stimuli used in Experiment 3.1. (contrast has been enhanced for reproduction). All patterns contain 64 elements and have a mean orientation of 90°. The standard deviations of the Gaussian distribution of orientations are (a) 4°, (b) 8°, (c) 16° and (d) 32°.

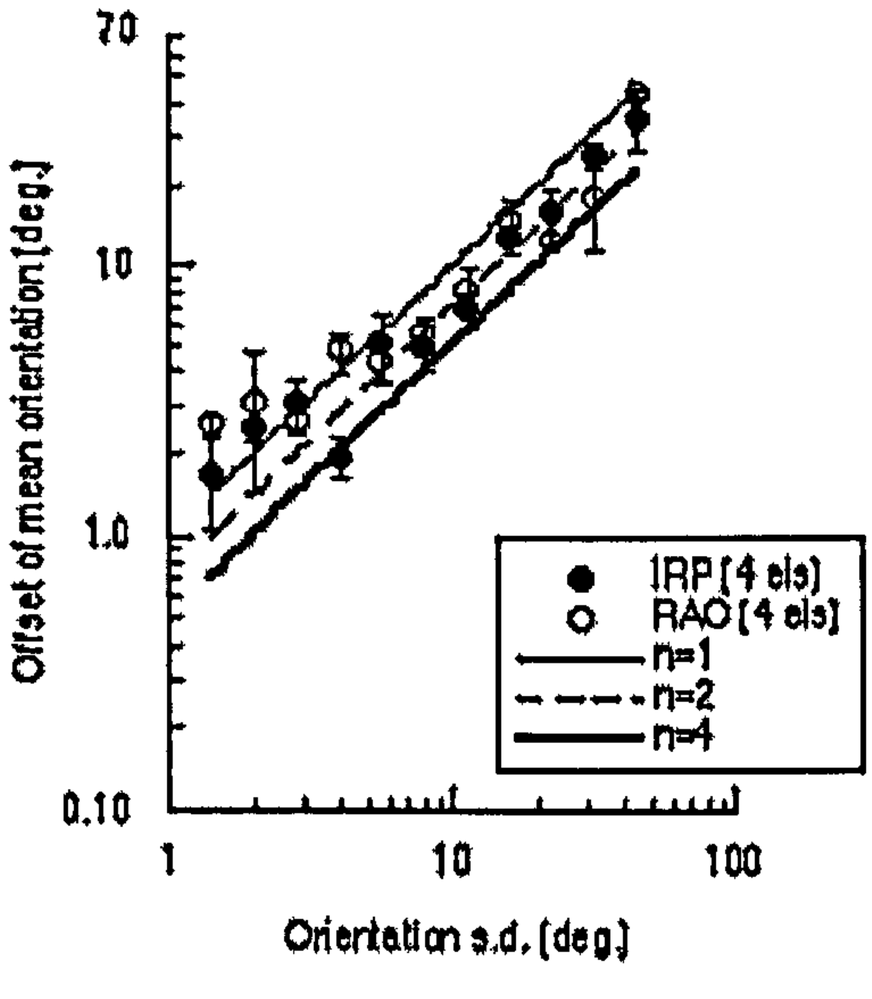
3.4 Experiment 3.1: Combining orientation measures

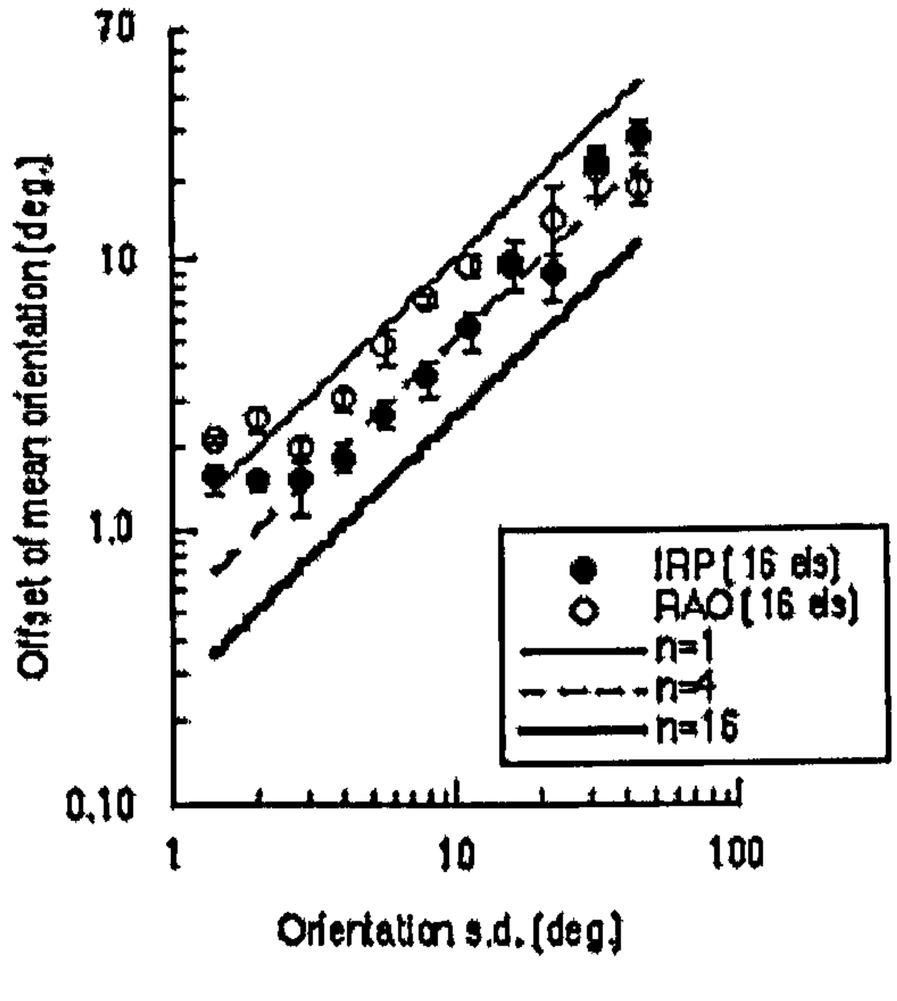
The first experiment was performed to establish whether the observers could combine multiple orientation measures to estimate the average orientation of a texture patch. To test this hypothesis a judgement was used which required subjects to make an estimate of average orientation. Subjects were presented with a set of oriented Difference-of-Gaussian patches whose orientation varied around some mean value. The task was to decide whether the mean orientation of the set was clockwise or anti-clockwise from vertical. The threshold offset of the mean orientation of the set required to perform the discrimination was measured as a function of the orientation variability of elements. Orientations of DoG patches were drawn from Gaussian random distributions. The standard deviation of this distribution was varied from 1.41° to 45° in half-octave steps and the experiment was run using patches containing 4, 16, 64 and 256 DoG patches. All textures were presented for 100 ms.

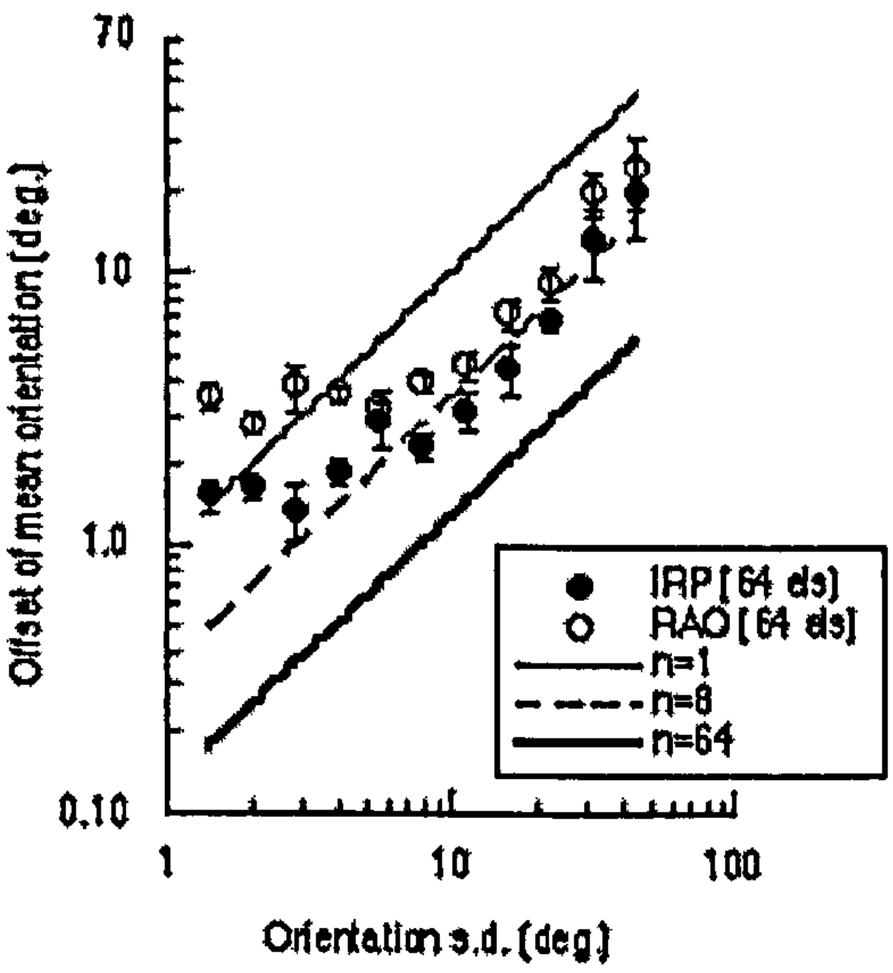
Figure 3.4 shows the results from this experiment for two observers. Assuming that there is no noise on one's estimation of the individual orientation of elements or in one's combination of those measures to form an estimate of the mean, then the only limit on one's performance in this task is the orientation variability of elements and the number of orientation estimates one uses. Given a set of n measures with standard deviation σ_{stim} then the error on an estimate of the mean will be:

$$\sigma_{estimate} = \frac{\sigma_{stim}}{\sqrt{n}}$$

Since the thresholds measured in this experiment directly estimate the standard deviation of the observers' estimate of the mean we can use this expression directly as an ideal estimator of mean orientation. Predictions from it, using different numbers of estimates, are superimposed on Figure 3.4. It is clear that subjects' performance systematically differs from this set of predictions, particularly at low levels of orientation variance. It seems that there is some degree of intrinsic noise on the subjects' estimates of the mean orientation which prevents them from achieving threshold less than around 1.5°. As the number of blobs in the patterns increases they can achieve this maximum level of performance with progressively higher levels of orientation variability. Because the model presented is an ideal observer, the number of estimates that it predicts subjects are using should only be taken as an absolute *minimum*. Any intrinsic noise may mean that more estimates are actually used.







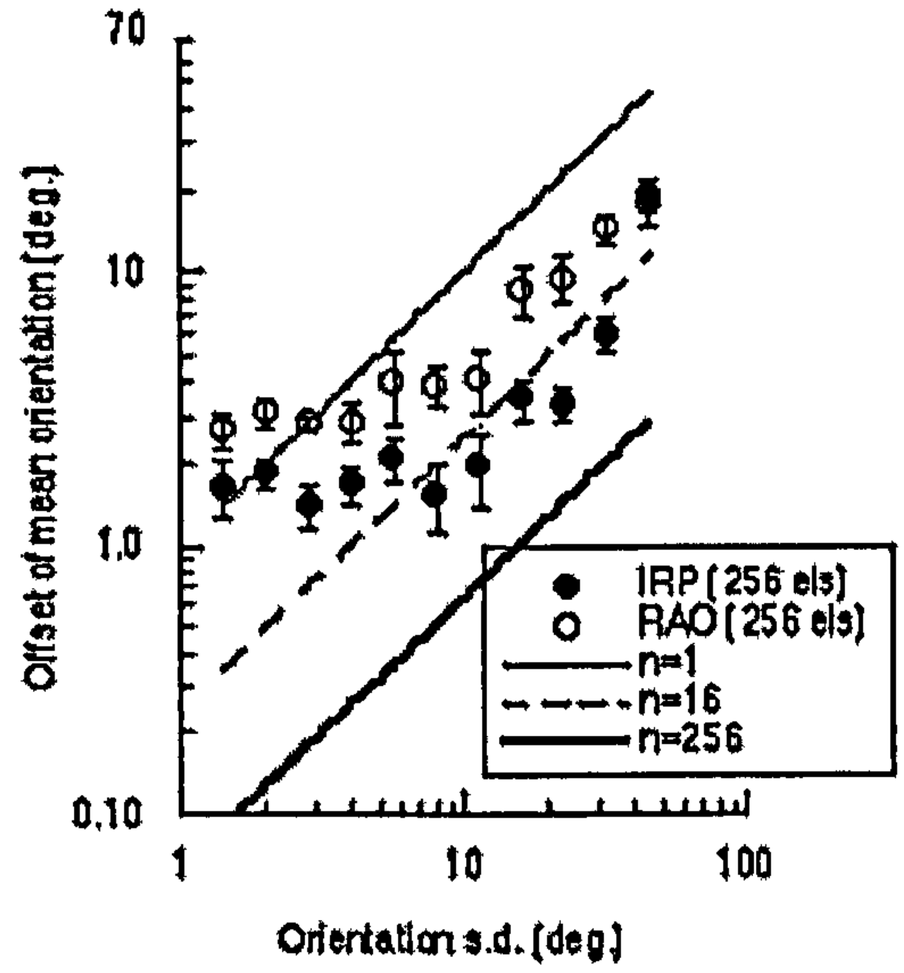


Figure 3.4. Results for two subjects from Experiment 3.1 with 4, 16, 64 and 256 texels. The abscissa represents the standard deviation of the Gaussian distribution from which orientations were drawn. The ordinate is the threshold offset of the mean orientation from vertical (in degrees) required to discriminate the texture from vertical. Predictions from a noise-free model averaging different numbers of estimates are also shown. (Abbreviations for figure: "els" (elements), "s.d." (standard deviation), "deg." (degrees)).

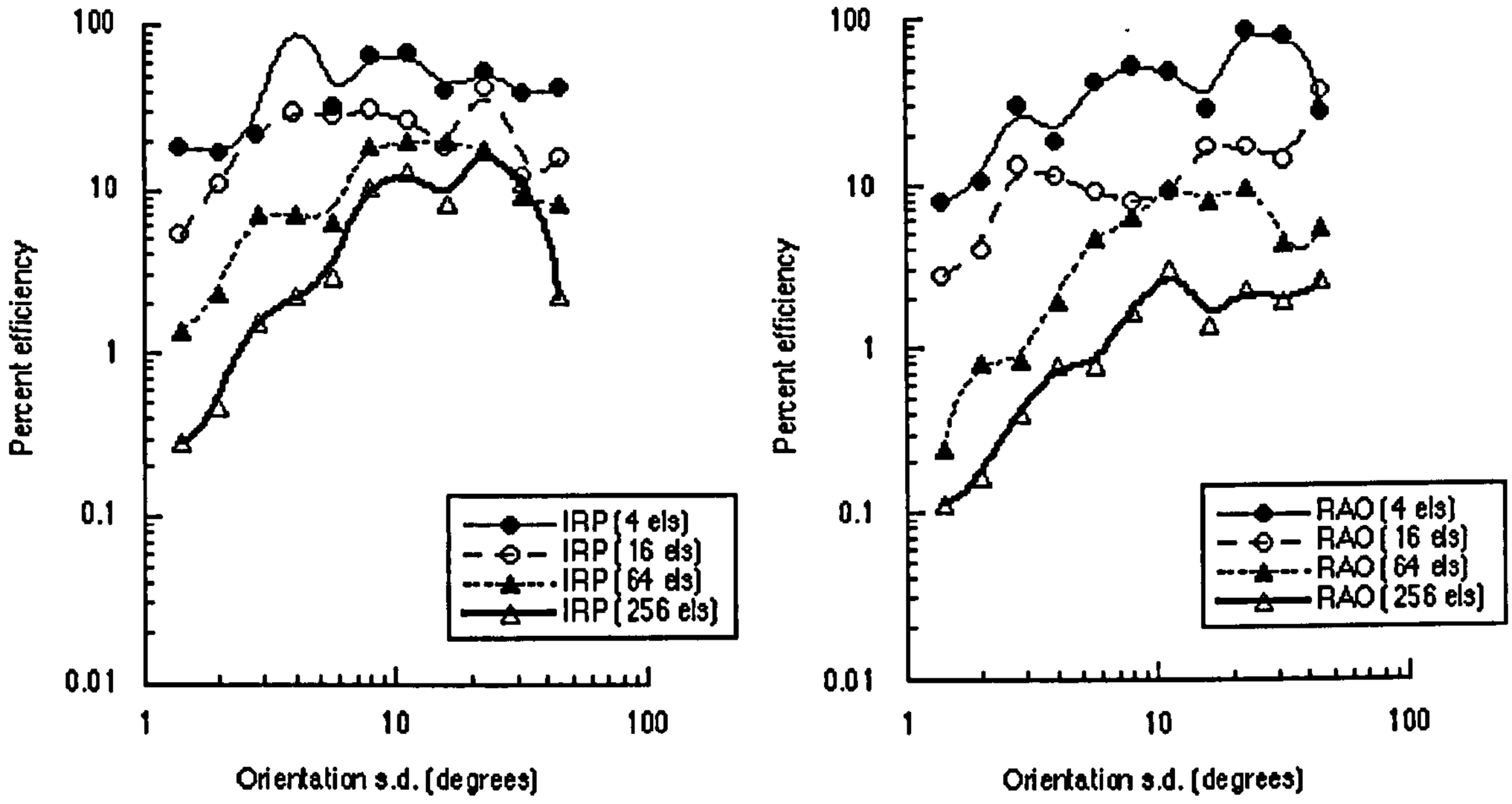


Figure 3.5. Results from subjects re-expressed as percentage efficiencies. Note decreasing levels of efficiency as the number of elements increases.

Figure 3.5 shows data re-expressed as percentage efficiencies. Given the original threshold (σ_{obs}), efficiency is calculated as:

$$F = \left(\frac{\sigma_{stim}}{\sigma_{obs}\sqrt{n}}\right)^2 \times 100$$

Subjects' efficiency drops steadily as the number of elements increases. Mean plateau efficiency, defined as the mean of the five rightmost data points from each condition, are as follows (for IRP and RAO respectively). (4 elements) 48.5% and 53.7%, (16 elements) 23.0% and 18.9%, (64 elements) 15.0% and 7.4%, (256 elements) 10.5% and 2.2%. Thus with small numbers of elements observers are highly efficient at this task.

The main point to note from these data is that subjects' performance always exceeds the predictions of the model when only using one measure of orientation. Given that this model is completely noise-free this strongly indicates that observers are combining individual orientations estimates to perform the task. The fact that subjects are performing at maximum levels of accuracy at higher levels of orientation variability in the 256 blobs condition than in the 64 blob condition suggests that the number of estimates used is still increasing above 64 elements.

In conclusion, this experiment has shown that in making estimates of average orientation subjects can integrate multiple orientation cues. The problem with this experiment

is that it will not separate the predictions from any number of models for estimating the central tendency of a set of measures: median, mode, etc. Because a Gaussian distribution of element orientations was used the mean, median, etc. all fall in the same location. The next set of experiments explicitly compared specific strategies for representing average orientation.

3.5 Experiment 3.2: Separating mixed texels sets

Subjects were sequentially presented with two stimuli (see Figure 3.6) each composed of two intermingled sets of texels. Sets contained n_1 and n_2 elements and had the same orientation variance ($\sigma = 6°$). The orientation of texels was determined using the sum of two Gaussian probability distributions with means of θ and $\theta + \delta\theta$. The total orientation probability distribution for each stimulus is then defined as:

$$p(\theta) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\left[p_1 \exp\left[-\frac{(\theta-\bar{\theta})^2}{2\sigma^2}\right] + (1-p_1)\exp\left[-\frac{(\theta-\bar{\theta}-\delta\theta)^2}{2\sigma^2}\right]\right] \qquad (3.5)$$

Where $p_1 = n_1/(n_1+n_2)$. Figure 3.6 shows examples of the stimuli used. In the reference stimulus the means of the two distributions were identical ($\delta\theta = 0°$). In the comparison stimulus a non-zero increment was added to one of the means. The threshold $\delta\theta$ was determined for the task of reporting which of the fields had orientations drawn from distinct distributions. Specifically, subjects were asked to decide which texture was composed of two sets of elements, with different overall orientations, from the texture which only contained one set. In all stimuli the standard deviation (σ) of the distributions was 6°. Two effects determined the choice of σ. Firstly, aliasing of the probability distribution due to pregeneration of DoGs in one degree steps, which becomes problematic at small values of σ. Secondly, the fact that a small number of elements poorly characterise a Gaussian distribution at very large values of σ.

The orientation of one set (θ) was randomised from trial to trial to eradicate remove mean orientation cues from the textures. The total number of texels, N, remained the same within a condition. Thresholds as defined above were determined at relative numerosities ranging from 1:1 to 1:64 evenly spaced on log axes. All subjects were run with 64 and 256 total texels.

Results are shown in Figure 3.7. Thresholds are lowest with equal number of texels drawn from both sets and increase with the relative difference in number. There is no apparent difference between the results from the conditions with 64 and 256 total elements.

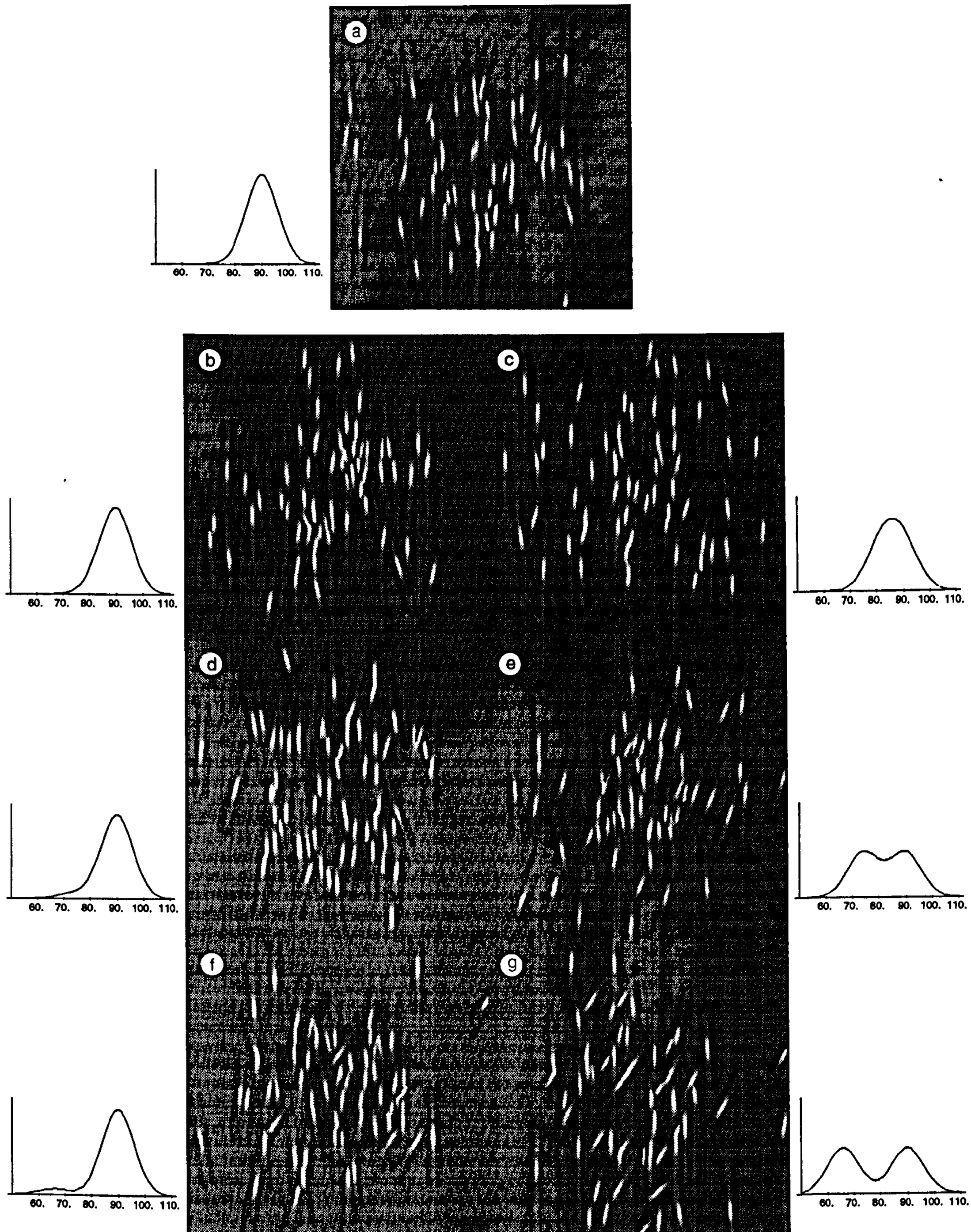


Figure 3.6. Example stimuli from Experiments 3.2-3.5 (contrast-enhanced for reproduction). All textures contain 64 elements and the orientation of the more numerous set is 90°. (a) shows the reference stimulus, containing two superimposed texel sets (i.e., the cue $\delta\theta = 0°$). (b,d,f) Two intermingled textures where one set has 16 times as many elements as the other ($n_1 = 16n_2$). The relative offset of the two sets, $\delta\theta$, is (b) 0°, (d) 8°, (f) 16°. (c,e,g) As above but for a texture where one set has twice as many elements as the other ($n_1 = 2n_2$). Relative offsets are (c) 0°, (e) 8°, (g) 16°. Graphs show the p.d.f.s for the adjacent texture.

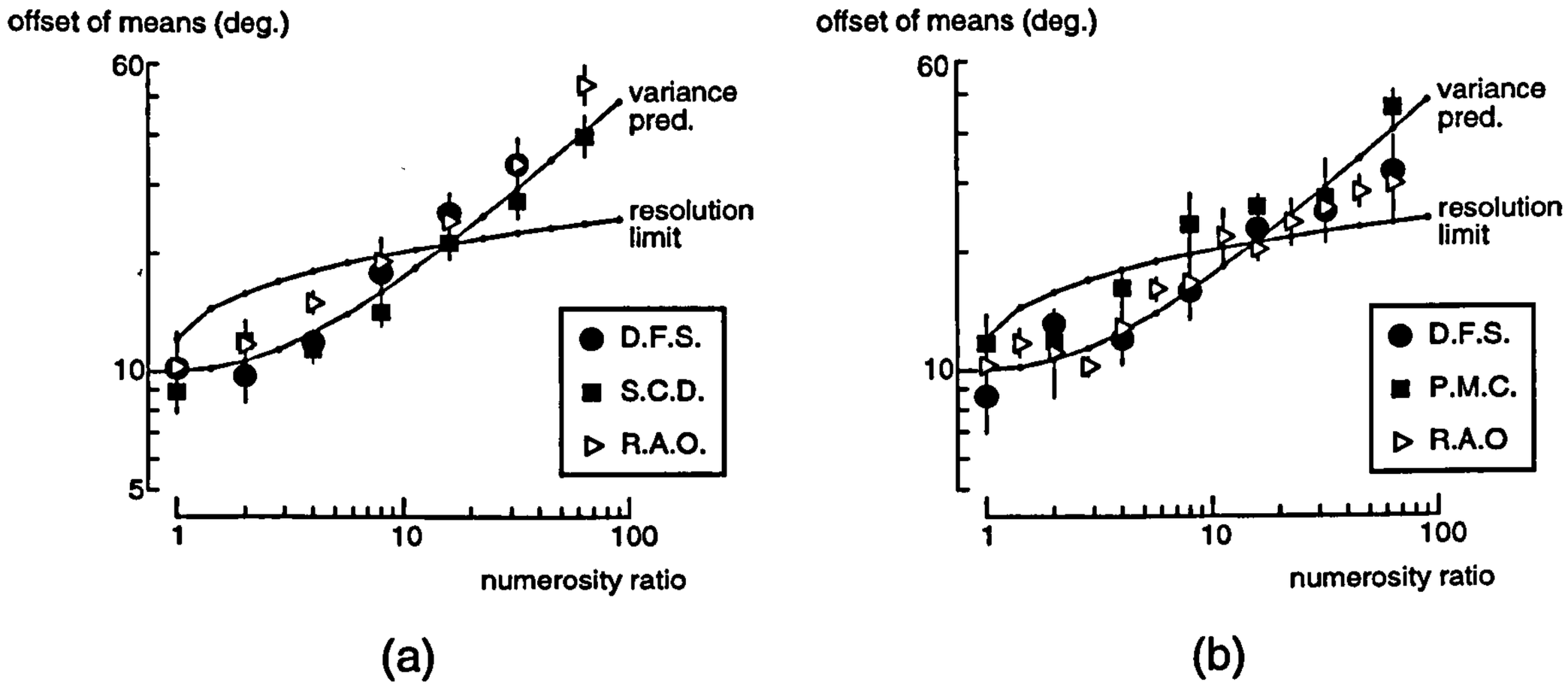


Figure 3.7. Results for three subjects from Experiment 3.2 with a total of (a) 64 and (b) 256 texels. The abscissa represents the relative number of texels in the two sets ($n_1/n2$); the ordinate represents the threshold offset of means in degrees. Predictions from a global variance judgement and the theoretical resolution limit of the distributions (as defined in the text) are also shown.

Figure 3.7 also shows the theoretical "resolution limit" for the stimuli employed in Experiment 3.2. In the spatio-luminance case, Watt and Morgan (1983a) define this as the point beyond which the number of stationary points in the first derivative of the luminance profile changes from one to three. If it is assumed that the observed distribution of element orientations is the same as the probability density functions used to generate them, then the orientation distribution of the patterns may be treated analogously to the spatial case and resolution limits calculated. For low ratios of texel number (<1:8) all subjects perform below this resolution limit for the two distributions. This is in agreement with the findings from Watt and Morgan (1983a). They suggested that because performance was not limited by resolution, the task amounted to one of *width discrimination*. Similarly the conclusion here is that subjects used some measure of the range of orientations in the texture patch (e.g. global variance) to perform the task at sub-resolution limit offsets.

To investigate this more thoroughly, the data were fitted with predictions derived from a variance discrimination task. In this task subjects reported which of two patches, each with a single set of texel orientations, had the greater variance. The reference set had orientations drawn from a distribution with $\sigma = 6°$. The cued set had a larger standard deviation. The threshold increase in standard deviation was determined for discrimination. Patches had randomised mean orientation: other experimental details were identical to those described in the Methods section (Section 3.3).

The threshold difference in standard deviation of orientation ($\delta\sigma$) was determined for each subject. The average threshold was 2.0°. The offset of distribution means producing this rise in *combined* distribution variance were calculated for each of the numerosity ratios from Experiment 3.2. The mean (θ_c) and variance (σ_c^2), of two combined distribution is given by:

$$\bar{\theta_c} = \frac{n_1\bar{\theta}_1 + n_2\bar{\theta}_2}{N} \tag{3.6}$$

$$\sigma_c^2 = n_1\sigma_1^2 + n_2\sigma_2^2 + n_1(\bar{\theta}_1 - \bar{\theta_c})^2 + n_2(\bar{\theta}_2 - \bar{\theta_c})^2 \tag{3.7}$$

where n_1 and n_2 are the number of elements in two sets, and N is the total number of elements ($n_1 + n_2$). Sets have means of θ_1 and θ_2 and standard deviations of σ_1 and σ_2. Given that the standard deviations of sets are equal, $\sigma = \sigma_1 = \sigma_2$ is substituted in Equation 3.7, and combined with Equation 3.5 to give:

$$|\theta_1 - \theta_2| = \frac{N^{3/2}\sqrt{\sigma_c^2 - \sigma^2}}{\sqrt{N^2 n_2 - 2N n_2^2 + n_1 n_2^2 + n_2^3}} \tag{3.8}$$

Substituting σ_c with the discrimination threshold standard deviation of a single set gives a prediction of the mean offset of two sets which produces this cue in standard deviation. These predictions are plotted for a variety of relative set numerosities in Figure 3.7. Note that no further fitting has been used. The variance predictions provide a good fit to the data. This is evidence that subjects use global variance information to perform this task.

The analysis above assumes that it is the *relative* number of elements in each set which determines subjects' performance on the task. To determine whether this is correct, a control condition was run. In the stimuli used, each set contained equal numbers but the total number varied from 16-512 texels. Data in Figure 3.8 show that within the range tested, total number has no apparent effect on performance. For all subsequent experiments a total of 64 texture elements was employed and pilot studies indicated that there was little difference in performance on any of the tasks when using 256 elements.

To summarise the data from Experiment 3.2: the discrimination of a texture with orientations drawn from a single set from a texture with orientations drawn from two sets with separate mean orientations, does not necessarily require resolution of component distributions. Fits to the data suggest that the task being performed is a judgement of

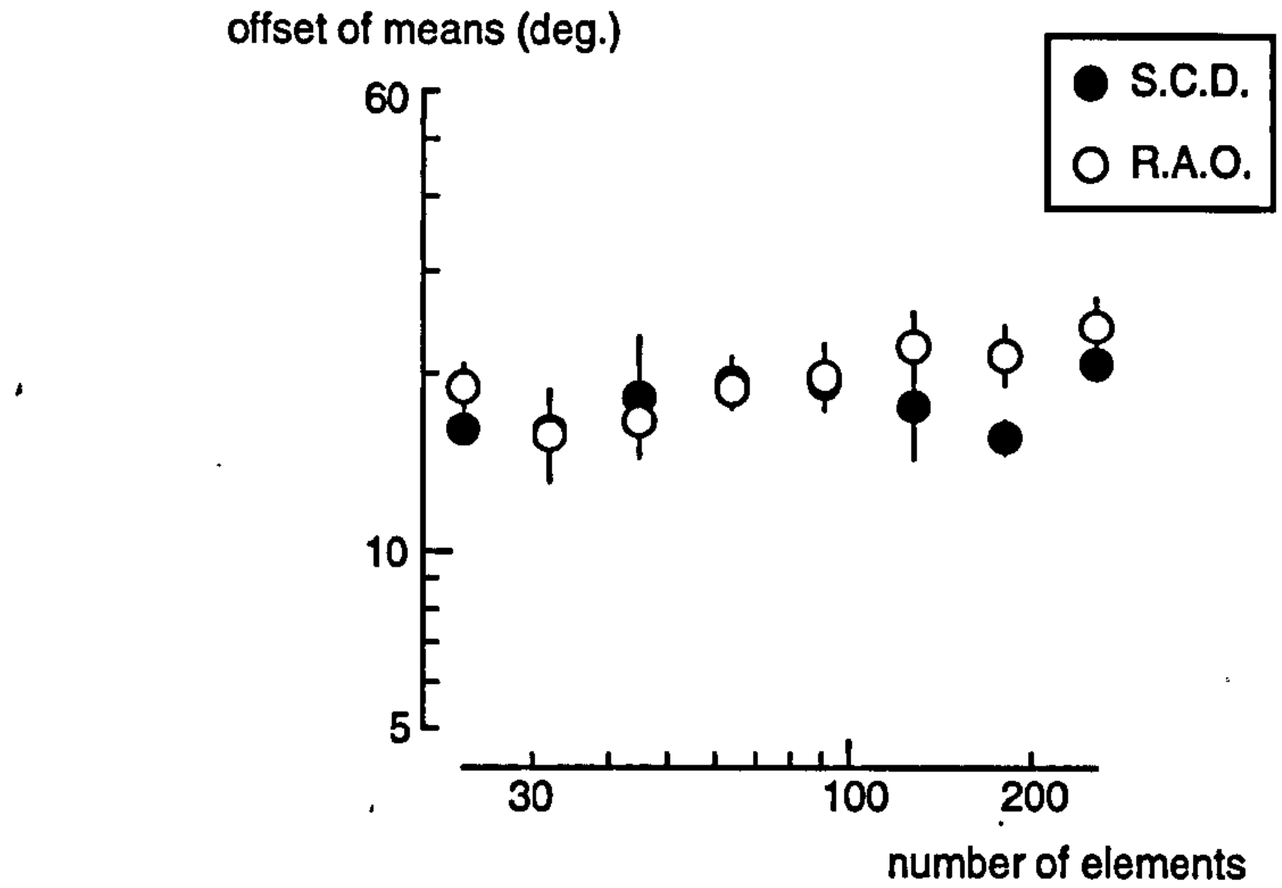


Figure 3.8. (a)-(c) The effect of the total number of texels on the task used in Experiment 3.2. The abscissa represents the total number of texels; the ordinate represents the threshold separation of the set means (in degrees) for discrimination of textures composed of two texel sets from those composed of one.

the global variance of the orientation. This result holds over a range of total numbers of texels.

3.6 Experiment 3.3: Judging the relative orientations of mixed texel sets

A texture patch composed of two sets containing unequal numbers of texels was presented. Subjects reported whether the less numerous set of DoGs was clockwise or anti-clockwise to the other set. This is a *numerosity asymmetry* task. This task used a single interval and to prevent the subjects exploiting the mean of the distributions as a cue, the mean orientation of one set was randomised over a 180° range. The experimental details were identical to those given in the Methods section (Section 3.3) with the exception that a longer exposure duration of 500 ms. was used, since pilot trials indicated that subjects were unable to obtain reliable thresholds at shorter exposure times. Subjects were given considerable practice on this task before data collection.

The results (shown in Figure 3.9a) showed that subjects appear to require resolution of distributions to perform the asymmetry task. This suggests that relative numerosity information about distributions is only derived if those distributions are distinct *perceptual* entities separated by at least the resolution limit. It also shows that the direction of skew of the orientation information is not available to subjects.

Comparison with the last experiment is difficult. In Experiment 3.3 subjects deter-

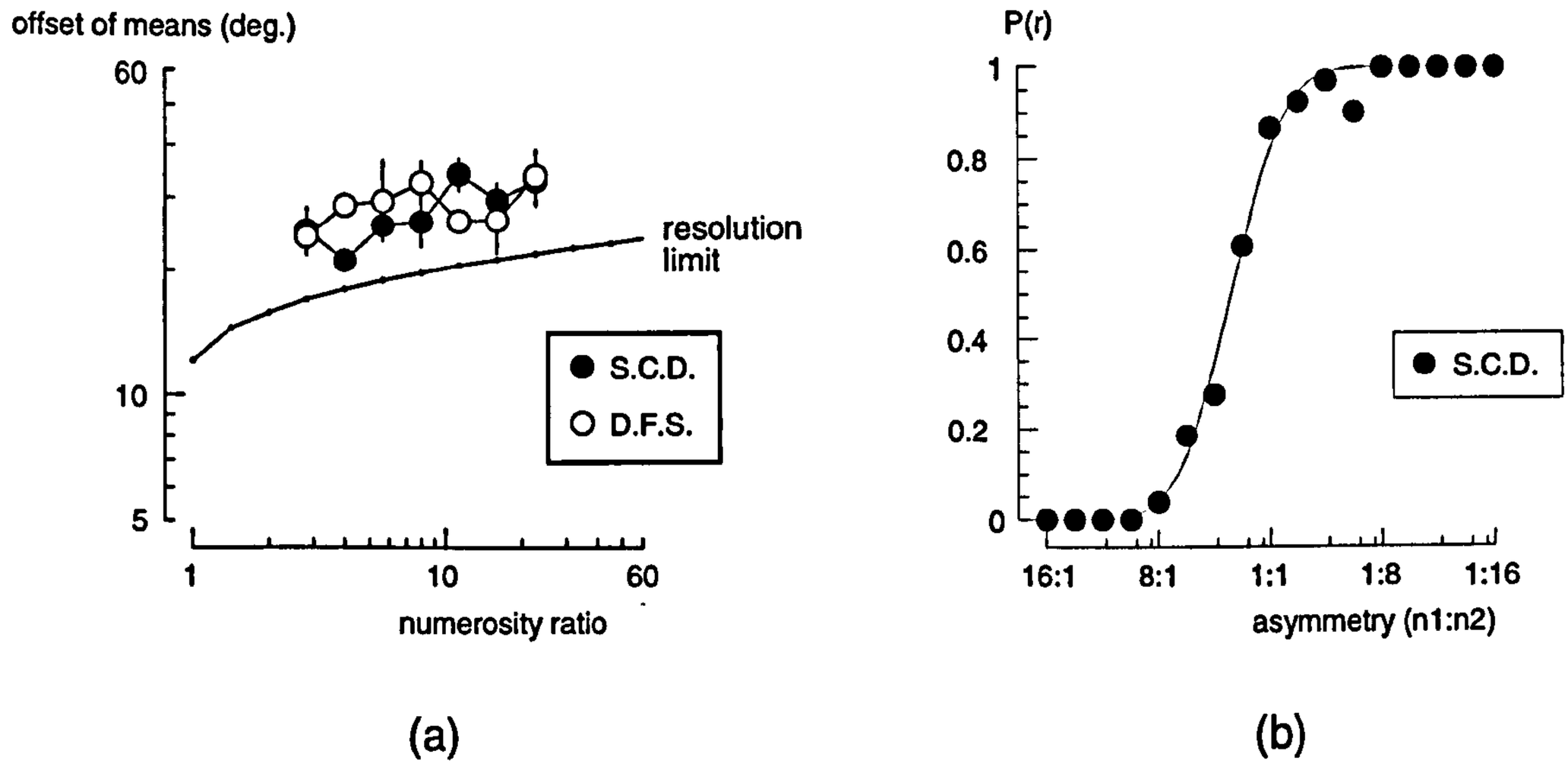


Figure 3.9. (a) Threshold offsets for comparison of the orientations of two sets containing different numbers of texels, as a function of relative number. Note that all thresholds exceed the theoretical resolution limit. (b) Psychometric function for relative numerosity judgement with a fixed offset of set means ($\delta\theta = 45°$).

mined which set was which (according to the number of texels in each) and judged the mean orientation of one in relation to the other. The fact that performance in this task approaches the resolution limit suggests that elevation of thresholds, compared to Experiment 3.2, *is* due to the offset of the means of the two orientation distributions. However, it is possible that the increased complexity of the numerosity judgement was responsible for poorer performance. To confirm that it is the offset of mean orientations which determine performance, a control experiment was performed. The task was a numerosity judgement at a fixed offset of mean orientations ($\delta\theta = 45°$), for different ratios of numerosity. This experiment will show the subjects' performance as a function of the difference in number between the two sets but in the absence of any interference due to orientation.

Results as a function of relative numerosity of sets are shown in Figure 3.9b, where a psychometric function for the author is given. The abscissa is the ratio of the number of elements in the two sets. Asymmetries from 16:1 to 1.4:1 indicate that the mean orientation of the set with more elements was *clockwise* of the other set. Asymmetries from 1:1.4 to 1:16 indicate that the mean orientation of the set with more elements was *anti-clockwise* of the other set. Data from this control were collected using a method of constant stimuli (1024 presentations) and fitted with a cumulative Gaussian. It is clear that any effects of absolute difficulty of the numerosity judgement (in isolation from

orientation interference) should be insignificant for ratios greater than 1:4 (the standard deviation of the psychometric function shown in Figure 3.9b is 1 : 2.1).

3.7 Experiments 3.4 and 3.5: Judging the overall orientations of mixed texel sets

The next two experiments looked at the extraction of average orientation and tested the predictions of a number of localisation models, adapted from the spatio-luminance domain.

Subjects were presented with one texture field and asked to indicate whether the average orientation was clockwise or anti-clockwise relative to the vertical. In Experiment 3.4 the field was composed of two sets with mean orientations of $\bar{\theta_1} = 90°$ and $\bar{\theta_2} = 90° + \delta\theta$ respectively. The offset of the means of the two sets ($\delta\theta$) was systematically varied. Additionally thresholds were determined for comparing the mean orientation of a single set with vertical. All textures were presented for 100 ms.

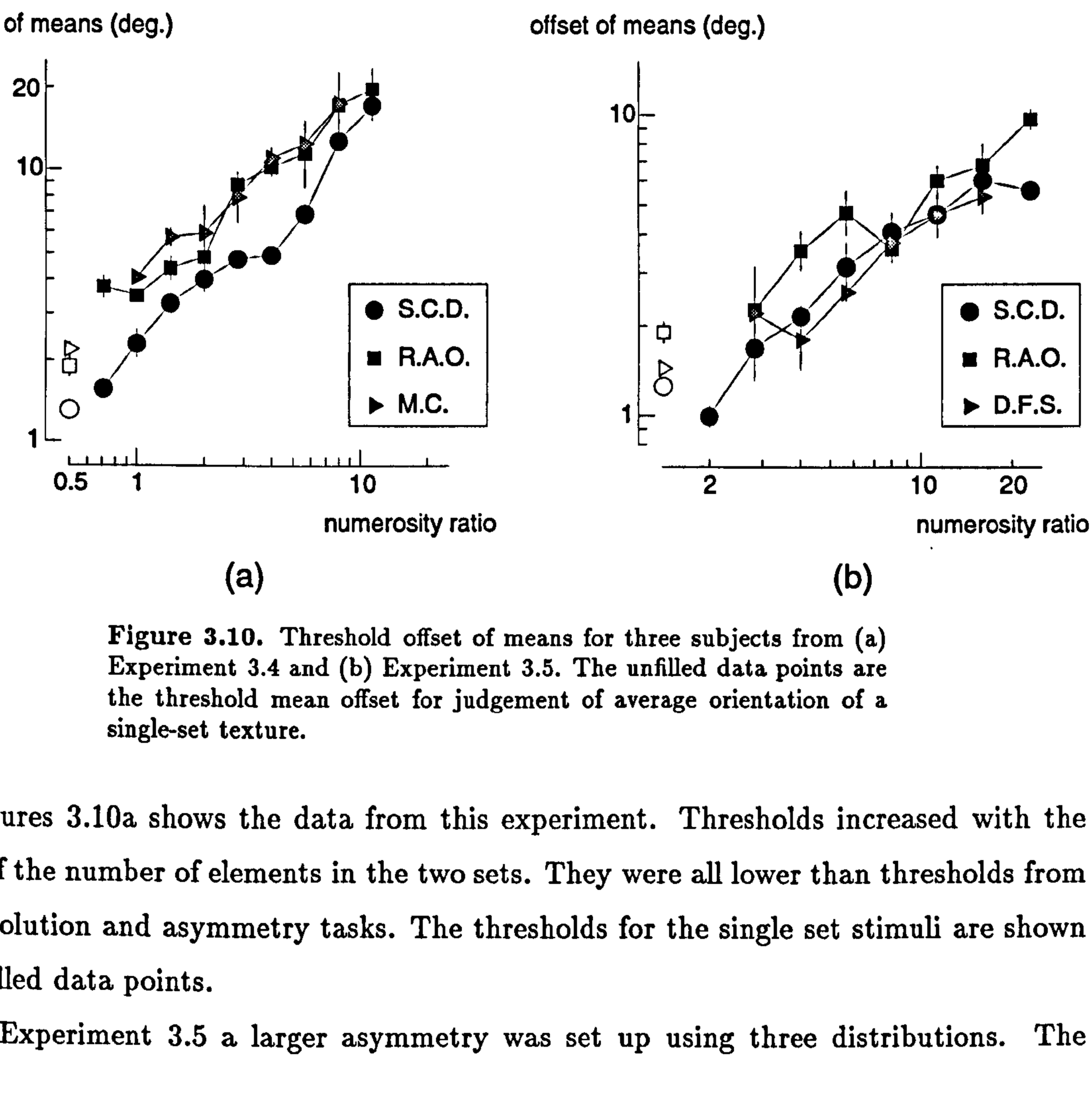


Figure 3.10. Threshold offset of means for three subjects from (a) Experiment 3.4 and (b) Experiment 3.5. The unfilled data points are the threshold mean offset for judgement of average orientation of a single-set texture.

Figures 3.10a shows the data from this experiment. Thresholds increased with the ratio of the number of elements in the two sets. They were all lower than thresholds from the resolution and asymmetry tasks. The thresholds for the single set stimuli are shown as unfilled data points.

In Experiment 3.5 a larger asymmetry was set up using three distributions. The

probability of a particular orientation was:

$$p(\theta) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma}\left[p_1 \exp\left[-\frac{(\theta-\bar{\theta})^2}{2\sigma^2}\right] + p_2 \exp\left[-\frac{(\theta-\bar{\theta}-\delta\theta)^2}{2\sigma^2}\right] + p_2 \exp\left[-\frac{(\theta-\bar{\theta}-3\delta\theta)^2}{2\sigma^2}\right]\right]$$

where $p_2 = (1-p_1)/2$ and all other variables are those given for Equation 3.5. Again the threshold $\delta\theta$ is for the discrimination of average orientation of a texture from vertical.

Threshold mean orientation offsets are shown in Figure 3.10b. Note that the numerosity ratios shown are based on the ratio of the number of elements in *one* of the flanking distributions to the number in the more numerous set (p_1/p_2). Unfilled symbols show the thresholds for the orientation judgement of a single set, as described above. Thresholds show the same monotonic dependence on numerosity ratio and are again lower than thresholds from the asymmetry and resolution tasks.

3.8 Comparison of features for coding mean orientation

In this section four schemes for coding the central tendency of an orientation distribution are examined. Consider a compound distribution of texel orientations with a particular ratio of the number of elements in each set, and an offset of component sets equal to the threshold value for that stimulus (as determined from Experiments 3.4 and 3.5). Features of this distribution will typically be located at different orientations (e.g. the mean of the function may not coincide with the peak orientation). The difference between the orientation of the candidate feature and vertical is assumed to be the cue that the subject uses. The threshold shift to judge the average orientation of a *single* set provides a comparison. If subjects use a particular cue to judge average orientation of the compound textures, then the threshold shift of this cue should equal the average orientation threshold for the single set judgement.

The four candidate features tested were: peaks, threshold edges, and two based on the combination of a number of measurements (zero-crossings and centroids). All features were extracted from idealised orientation distributions.

- Zero-crossing: the zero crossing model assumes convolution of the idealised orientation distribution with a smoothed second differential operator (convolution is wrapped around the 180° range). This is similar to the "orientational filters", proposed by Keeble and Morgan (1993), as a mechanism for discriminating oriented textures. It is also similar to a component of Paradiso's (1988) model of orientation perception, where the output of cells within an orientation column are filtered with

a Difference-of-Gaussian filter. In this case a Laplacian-of-Gaussian operator was used, of the form:

$$\nabla^2 G(x,\sigma) = \frac{1}{\sigma^2}\left(\frac{x^2}{\sigma^2} - 1\right) e^{-x^2/2\sigma^2}$$

Location was then defined as the mid-point between the outermost two zero-crossings of the convolution about vertical. The width of filter employed was the value minimising the Chi-square of the fit of the model to the data from Experiment 3.4. This value was a filter standard deviation of around 6° (mean=6.10°, s.d.=0.81°)

- Threshold edge: two points in the distribution were located which were equal to some threshold value. Location was defined as the mid-point between these points. The threshold employed again minimised the Chi-square of the fit of the model to the data.

- Peak: the location of a distribution is assigned to the maximum of the idealised orientation distribution.

- Centroid: the centroid was calculated using:

$$\bar{\theta} = \begin{cases} \theta_e & \text{if } -2\left[\sum_{i=1}^{180} p(\theta_i)\cos 2(\theta_i - \theta_e)\right] < 0 \\ \theta_e + 90° & \text{otherwise} \end{cases}$$

where:

$$\theta_e = \frac{1}{2}\tan^{-1}\left[\frac{\sum_{i=1}^{180} p(\theta_i)\sin 2\theta_i}{\sum_{i=1}^{180} p(\theta_i)\cos 2\theta_i}\right]$$

Note that integration occurs over the full range of orientations. In practice integration over a restricted range of orientations can be more appropriate (e.g. when a distribution contains two resolved entities).

Given a threshold difference in the mean orientations of the component sets, the size of the cue given by each of the candidate features may be calculated. These values should be equal to the size of the cue for the single-set texture at threshold offset. Cue sizes based on thresholds from Experiment 3.4 are shown in Figure 3.11a-c, and those based on thresholds from Experiment 3.5 are shown in Figure 3.11d-f.

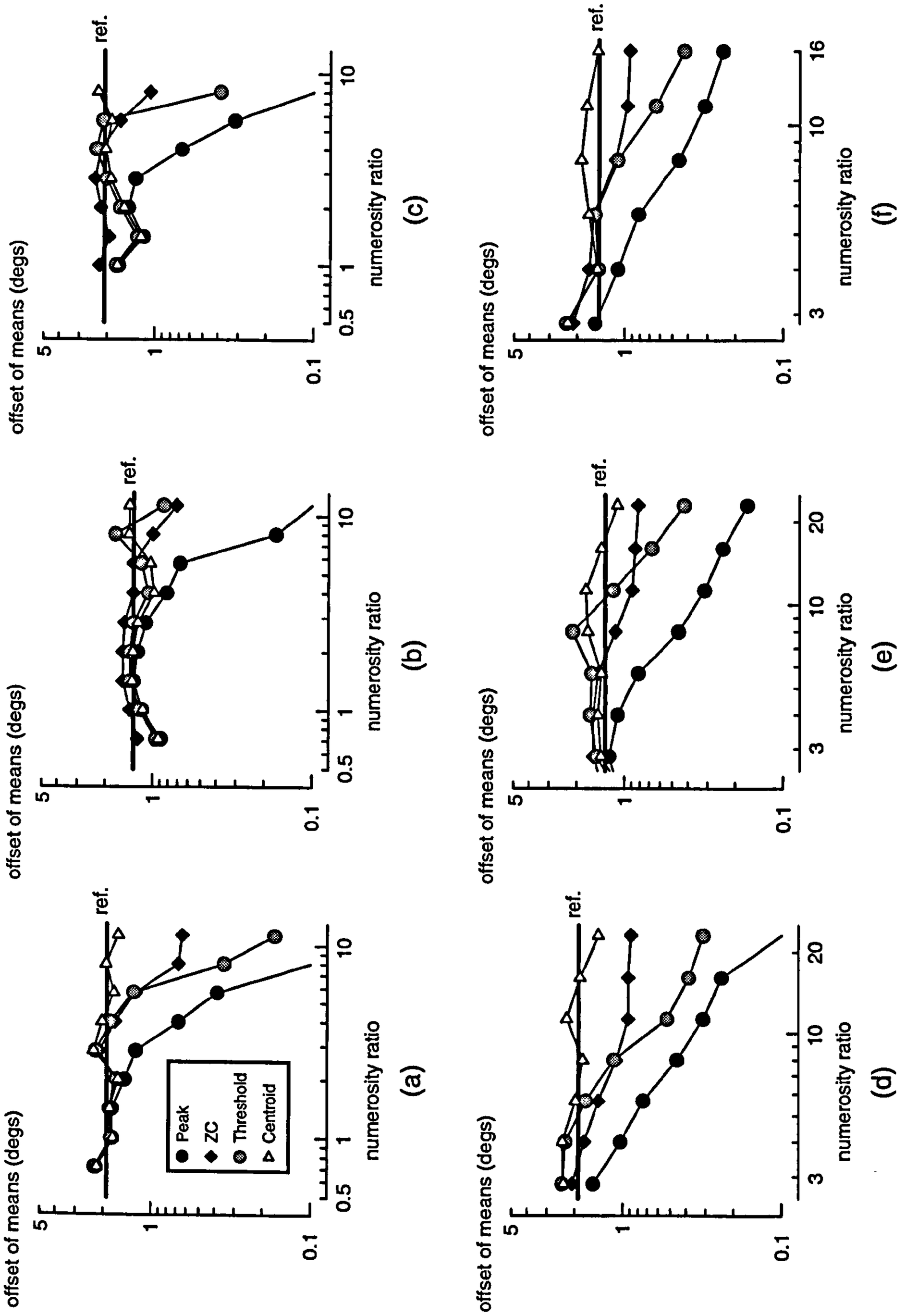


Figure 3.10. Subjects performance on the global orientation judgement re-expressed as shifts in different features in the idealised orientation distribution. (a-c) Data from Experiment 3.4 (RAO, SCD and GP, respectively) (d-f) Data from Experiment 3.5 (RAO,SCD and MC) respectively.

Figure 3.11.

		Threshold	Peak	Centroid	ZC
(Exp 3.4)	RAO	0.041	0.116	0.27	0.029
	SCD	0.047	0.168	0.057	0.019
	GP	0.168	0.324	0.181	0.018
(Exp 3.5)	RAO	0.423	0.897	0.051	0.24
	SCD	0.151	0.269	0.034	0.053
	MC	0.168	0.324	0.018	0.077

Table 3.1. Chi-squares for the fits of the four models tested using data from Experiments 3.4 and 3.5

Typically the models produce cues which are smaller than the threshold shift in central tendency, even given the idealised input, and can therefore be eliminated. The peak model in particular fails to explain the data (weighted Chi-squares appear tabulated in Table 3.1). The best fit is given by the centroid model although the threshold edge and zero-crossing model are also adequate. Note that the latter two models are fitted, whereas the centroid model has no free parameters. The Chi-squares given are all very low. The discrepancy between these values and the fact that, from examining the graphs, some models appear to fail badly, is due to the weighting used. Most models perform poorly at high numerosity ratios. The data from subjects at these points tend to be noisy so that the weighting of Chi-squares based on standard error will largely discount these data.

For the simulation of Experiment 3.5, the same threshold values and filters were used (except in the case of subject DFS where no data from Experiment 3.4 were available, and these values were derived using the data from this experiment). Figure 3.11d-f demonstrate that the increase in asymmetry of the stimuli in Experiment 3.5 further separates performance of the four models. Again the centroid model best explains the data and this is reflected in the very low Chi-squares, given in the lower half of Table 3.1.

These results suggest that the centroid of a set of orientation measures can be used to estimate dominant orientation. This model makes two primary assumptions: firstly, that orientation measures are always extracted from the whole image; secondly, that integration over the whole range of orientations should be used to make the estimate. These two assumptions are almost certainly false. The first would not allow for the perception of multiple orientations within a texture; the second would not allow for the perception of globally organised flows such as rotational fields. Experiment 3.6 examines how *local* estimates of orientation are made, both in the sense of spatial position and orientation

distributions. It examines a method for effectively extracting interesting features from orientation-frequency information.

3.9 Experiment 3.6: Limiting the orientation range

The simulations presented in this chapter suggest that average orientation is coded using some centroid measure. This centroid, for the purpose of modelling, integrated orientation measures over a 180° range. However, Experiment 3.3 demonstrated that subjects are quite capable of representing *multiple* orientations derived from a texel set. This suggests that a model capable of explaining both results must have some way of limiting the range of orientations over which the centroid is calculated.

The zero-crossing model presented in Section 3.8 practically limits the orientation range used. Orientation histograms are filtered using a one-dimensional Laplacian-of-Gaussian, a threshold is applied and zero-bounded regions of response are characterised by the mid-point between these ZCs. This model was not as good an explanation of subjects behaviour in Experiments 3.4-5 as the centroid model. However a practical way of increasing the accuracy of this model is to use the centroid of regions between zero-crossings. This model limits the orientation range effectively but retains the accuracy of the centroid estimate.

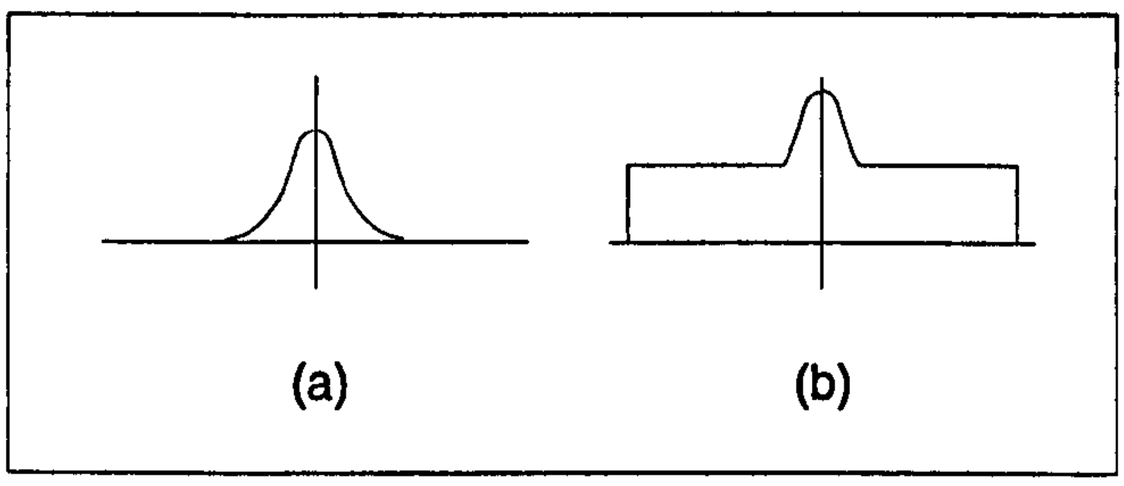


Figure 3.12. Orientation p.d.f.s (a) A Gaussian distribution. (b) The sum of a Gaussian and a uniform distribution.

How would one test such a model? An estimate of mean orientation from a system which could not restrict the range of orientations analysed, is affected by *all* local orientation estimates. Consider the orientation p.d.f.s shown in Figure 3.12: (a) shows a Gaussian distribution and (b) the sum of a Gaussian and a uniform distribution. Suppose one were interested in estimating the mean of the Gaussian orientation distribution, defined by a sub-population of the orientation estimates. A system averaging across all orientations would be much more affected by the background, uniform distribution than

a system which could restrict analysis to the region of the Gaussian. The ZC model, for example, will effectively suppress the d.c. component of Figure 3.12b.

The following experiment used textures with orientations similar to Figure 3.12 to examine how the estimation of the mean orientation of a set of similarly oriented elements was affected by the number of uniformly oriented elements that the set was mixed in with.

Unlike previous experiments, the textures used here were generated by locating oriented patches on a perturbed grid, rather than allowing them to fall anywhere within the pattern. This was not only to ensure the visibility of all elements, but was also to allow the investigation of the importance of subjects' knowledge of the extent of the target texture. As well as presenting the cued texture on a uniform grey background, a second condition embedded the cued texture in a large field of uniformly oriented elements. This condition will be disruptive only if the subject cannot voluntarily restrict the region over which orientation estimates are averaged.

3.10 Methods

All apparatus and viewing conditions were identical to those described in Section 3.3.

3.10.1 Subjects

The author and two other observers, who were both naive to the purposes of the experiments, served as subjects. All were experienced psychophysical observers. SJE and SCD are corrected-to-normal myopics.

3.10.2 Stimuli

Textures were composed of small patches with luminance modulated by a Difference of Gaussian function, in order to restrict information to a narrow range of spatial scales (expressions for the DoGs are given in Equations 3.2-3.4). Elements were located on a perturbed grid. The mean separation of points on the grid was 1.25 degrees min. To avoid a number of elements co-aligning, the x and y positions of patches were allowed to vary $\pm 25\%$ of the mean separation. The overall size of the image was 512 pixels square which subtended 20.0 degrees square.

Unlike the experiments in Chapter 3, normalisation of the textures was not performed: because of the spatial distribution of the elements on a grid, overlap of elements could not produce values outside of the range of permissible grey-levels.

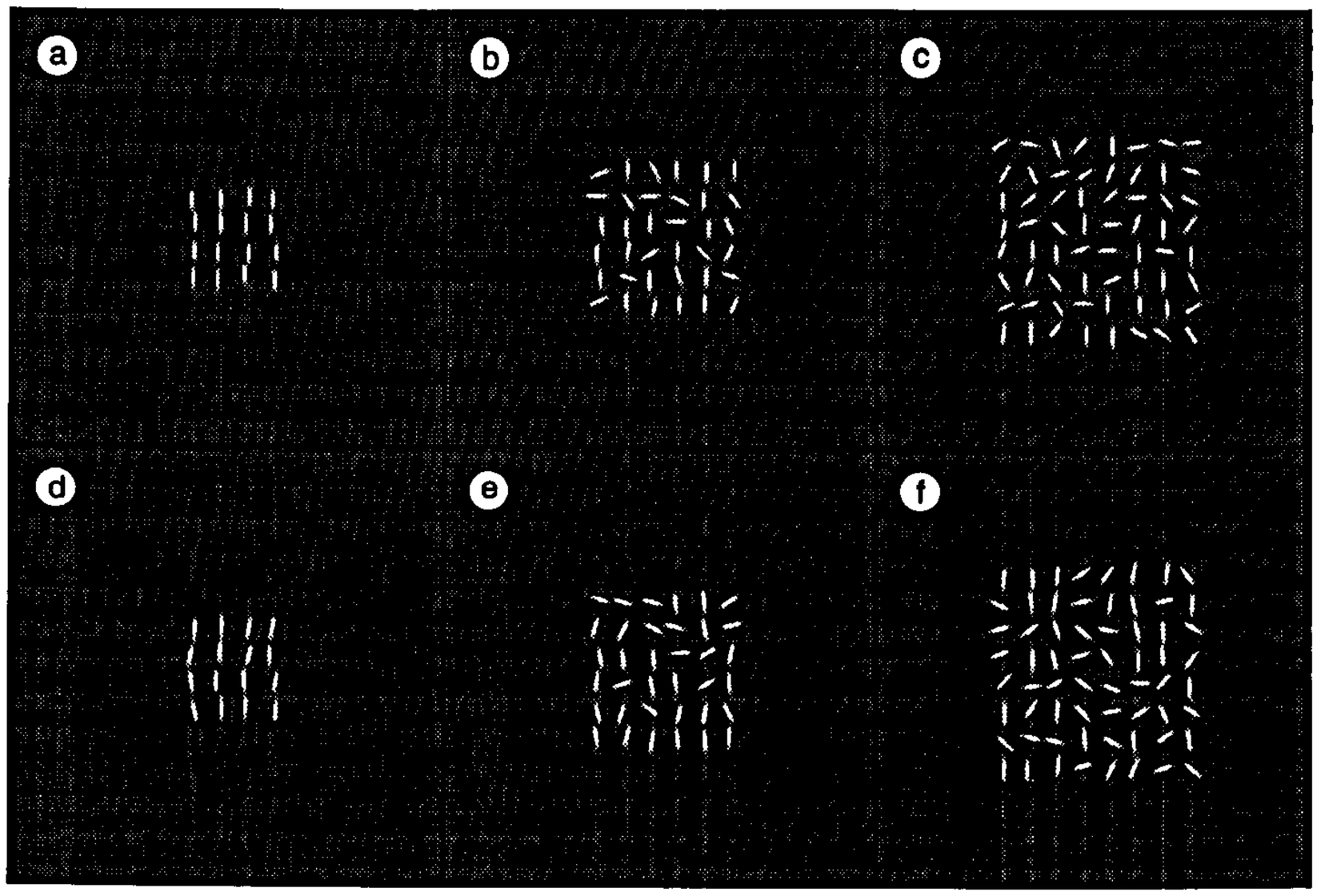


Figure 3.13. Example stimuli from Experiment 3.6. Sixteen elements, with a mean orientation of 90° are presented on a 4 X 4 grid (a,d), a 6 X 6 grid (b,e) and an 8 X 8 grid (c,f). The vacant positions are filled with randomly oriented elements. (a-c) Cued elements have no orientation variance. (d-f) Element orientations are Gaussian distributed with s.d.=8°.

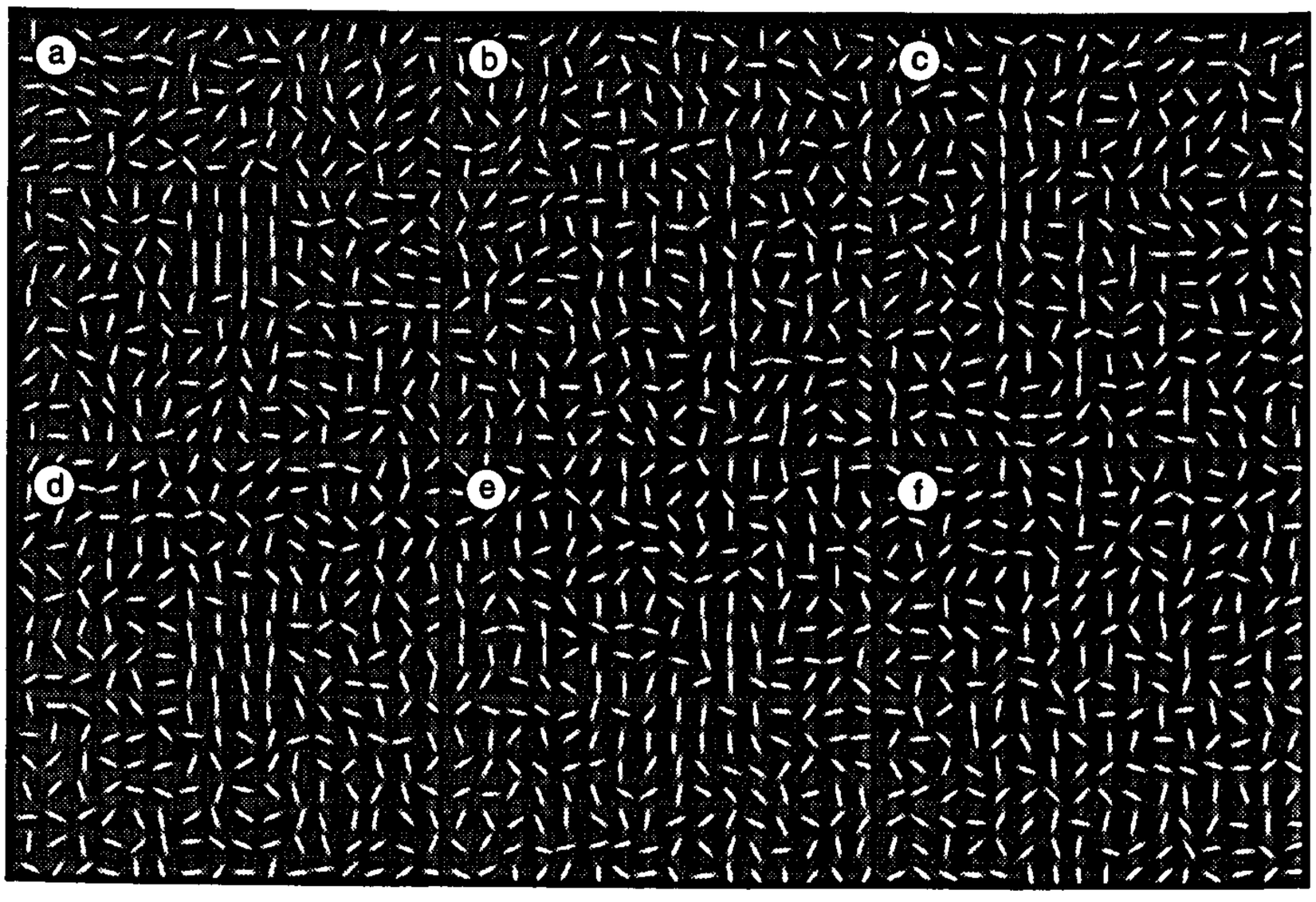


Figure 3.14. As Figure 3.13, except patches are all embedded in a 16 X 16 grid of randomly oriented features.

Each texture contained a population of sixteen elements drawn from a single Gaussian orientation distribution (the *target set*). These elements were arranged on a grid, the width of which was the independent variable. As the grid width increased from its minimum size (4×4), vacant positions were filled with randomly oriented elements (for examples see Figure 3.13). In the first (unembedded) condition, these grids were presented to subjects in the centre of an image with uniform grey-level. In the second (embedded) condition, these grids were placed surrounded by randomly oriented elements (Figure 3.14). The dimensions of of the background grid were 16×16.

3.10.3 Procedure

A single-interval judgement of mean orientation was used. Subjects were presented with a stimulus and asked to judge if the mean orientation of the target set was clock-wise or anti-clockwise from vertical. Textures were presented for 100 ms. in the centre of the display. An adaptive method was used to sample a range of offsets of the mean orientation, and all other details of data collection are similar to those presented in Section 3.3.

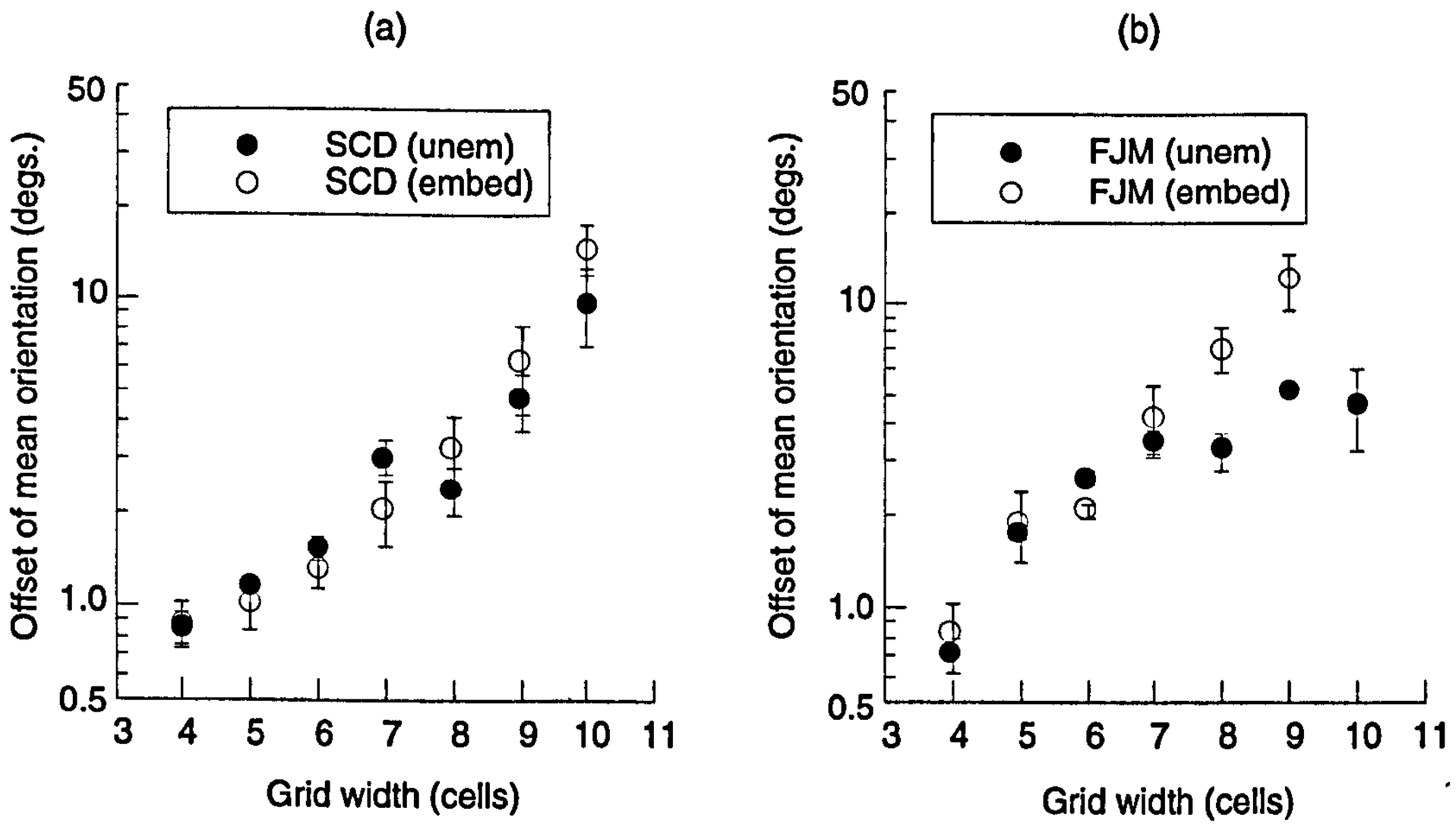


Figure 3.15. Results from two subjects for Experiment 3.6. Note that there is little difference between the embedded and unembedded patch condition.

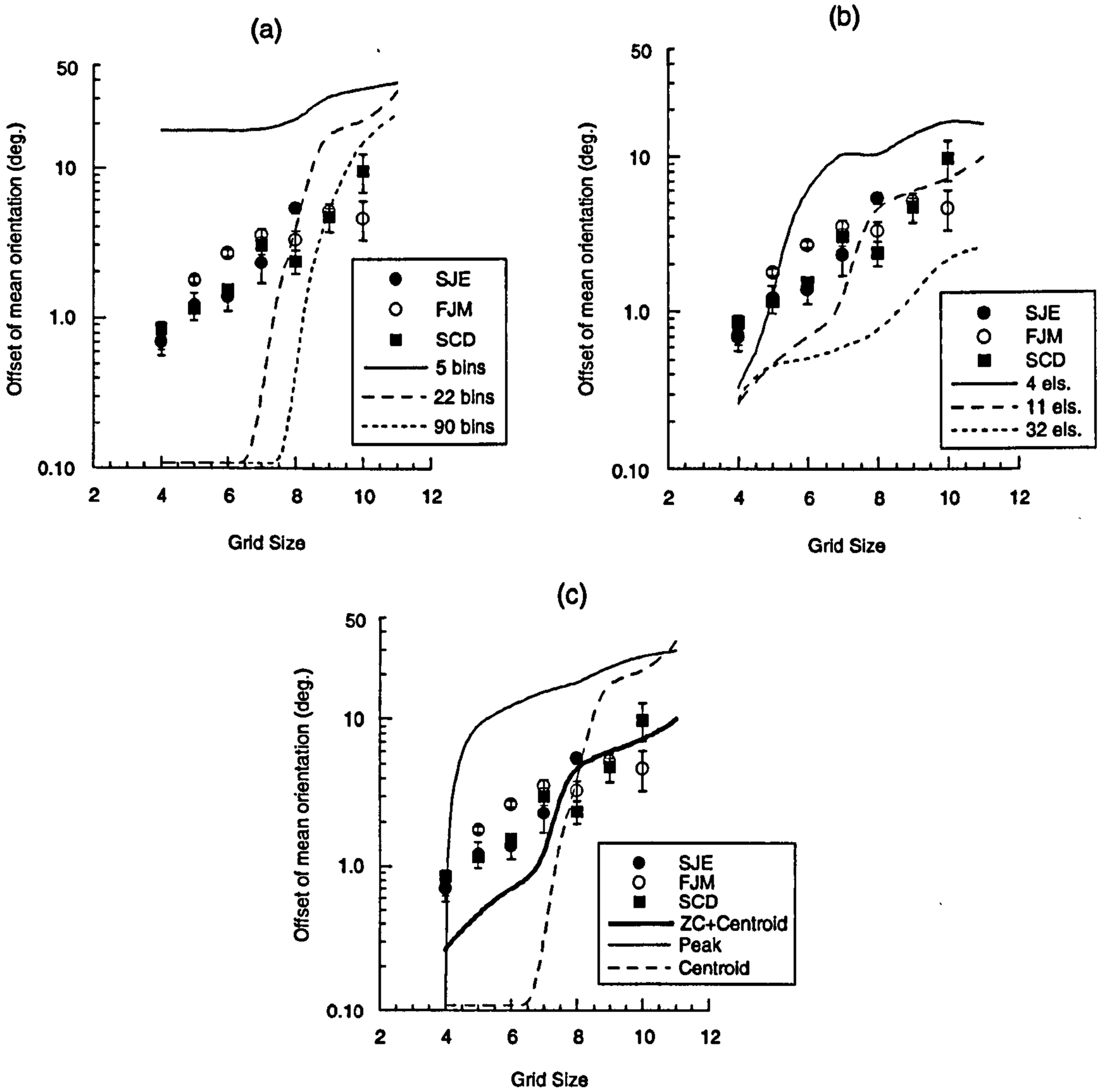


Figure 3.16. (a) Predictions from the peak orientation model, for the task from Experiment 3.6, as a function of the number of of bins used to from the orientation histogram. None produce good fits to the human data (large symbols), (b) Predictions from the ZC + centroid model, as a function of the number of estimates of orientation used. Note the gradual deterioration in performance as grid size increases. (c) The best fitting predictions of the three models to the human data.

3.10.4 Results

Figure 3.15 shows data from two subjects on this task. As patch size increases the task becomes progressively harder: performance degrades uniformly up to a cell size of around 10 by 10, beyond which point it is difficult to determine a threshold. The embedding of the patch in background of noise makes little difference to performance. This suggests that subjects have accurate control over the region of integration they use for estimating mean orientation.

3.10.5 Models

The data from this experiment were modelled using three systems for extracting average orientation. All of these systems operated on an idealised set of orientation measures, i.e. models did not use measurements made directly from images but instead operated on sets of orientation values which were calculated in the same way as for the texture stimuli. For all models a histogram was first derived from the orientation data. In the case of the peak model, the dominant orientation was taken to be the bin containing most elements. In the case of the centroid model a centroid was calculated using the procedure described in Section 3.8. For the ZC+centroid model the orientation histogram was convolved with a Laplacian-of-Gaussian filter, the scale of which (s.d.=6°) was derived from the simulations of Experiments 3.4 and 3.5. The response was thresholded at a value of one standard deviation from mean response and the zero-bounded region with the largest mass extracted. The centroid of this region of response was then calculated in the same way as described in Section 3.8.

For the peak model the number of bins, and the number of estimates used from the data were free parameters. For the centroid and ZC+centroid model, the number of estimates was a free parameter. The parameter determining the number of estimates set the *maximum* number of elements that could be averaged (a range from 4-64 was tested, for both models). Parameters were selected to minimise the chi-square deviation of the model from the data.

For each of the grid sizes tested on human subjects, 3 sets of 64 orientations were generated with the same signal to noise ratio (SNR) used in the experiment. The standard deviation of the estimated mean was calculated for each model operating on each each set. The three standard deviations from each model were averaged to give the mean predicted response of the different models at each grid size.

	Peak	Centroid	ZC + Centroid
SJE	6.88	247.3	2.56
FJM	8.20	272.4	2.30*
SCD	10.95	291.3	2.10

Table 3.2. χ^2 of the fits of the four models to data from Experiment 3.6.

Figure 3.16a shows predictions from the peak model as a function of the number of bins used to histogram orientation data. Changing the number of estimates used by the model and the number of histogram bins have very similar effects on the predictions on the peak model, and for that reason predictions are only shown for the model using 32 elements. With large numbers of bins, the peak model performs perfectly[2] up to a critical grid size. Beyond this point, performance collapses and is worse than human subjects at larger grid sizes. Decreasing the number of bins (or estimates) does not steadily shift performance up on this graph. At a critical number of bins, the model's performance suddenly becomes very poor (e.g. the performance with 5 bins shown in Figure 3.16a). Even given two free parameters this model does not produce a good fit to the data.

Predictions of the ZC + centroid model are shown in Figure 3.16b, as a function of the number of orientation estimates used. Note that the performance of this model smoothly deteriorates as a function of the grid size, showing the same trend as the human subjects.

Figure 3.16c shows the predictions of the three models for selecting average orientation which most closely match subjects' data. The predictions from the centroid models are uniformly worse than human observers and for that reason only the performance of the model using the maximum number of orientation estimates is plotted. By integrating over the entire set of orientations present in the texture, the model's performance drops rapidly as grid size and the amount of noise elements increases. The zero-crossing/centroid model, using 11 orientation estimates, seems to produce a good fit to the data. Performance drops off steadily as patch size increases. Subjects cannot completely ignore background noise, as the peak model can, but instead average over a limited range of orientations. This is confirmed by the Chi-square values for the goodness-of-fit of the models to each subject's data, which are tabulated in Table 3.2. For all three subjects, the goodness-of-fit measure of the ZC+centroid is much lower than for other fits.

[2]Perfect performance is plotted as a threshold of 0.1° to allow presentation on log axes

This experiment suggests that subjects can extract interesting orientation structure from noisy distributions of orientation information. Subjects' performance suggests that the mechanism for doing this cannot rely on the peak of orientations present or the simple centroid. Instead a system using a filtering operation to extract *relative* orientation frequencies (i.e. a method for removing the d.c. component of the signal) explains performance well.

3.11 Conclusions

There are six key findings from the experiments described in this chapter:

- Subjects can combine separate measures of orientation to form estimates of overall orientation (Experiment 3.1).

- Subjects can discriminate stimuli composed of two oriented texel sets from a field composed of only one set at mean orientation offsets which depend on the *relative* number of elements in each set. (Experiment 3.2).

- The global variance of a stimulus is the cue for the discrimination of textures composed of sets with equal mean orientation from those with sets with different mean orientation. Threshold orientation offsets can be below the theoretical resolution limit (Experiment 3.2).

- When the variance cue is eliminated by using a relative numerosity judgement, thresholds are higher and reliably above the resolution limit. Information about the skew of orientation distribution is not available. The resolution limit gives a lower limit on when texel sets become separate, in the sense that properties of those sets can be measured independently (Experiments 3.3).

- Assuming an idealised orientation distribution, a comparison of different features for representing central tendency indicates that a centroid model is the most appropriate of the models tested. (Experiment 3.4 and 3.5).

- Subjects do not necessarily use the centroid of measures taken over the whole orientation range but can calculate the centroid over a limited range of orientations, thereby limiting the effects of noise. (Experiment 3.6)

In Experiment 3.3 subjects clearly have access to relative numerosity information about two texel sets and an orientation primitive based on some portion of the orientation distribution is suggested. Although the zero-crossing model produces a reasonable fit to the data from Experiments 3.4 and 3.5 it was also run (with identical filter characteristics) on the data from Experiment 3.2 to yield predictions of variance discrimination thresholds. The model produced adequate predictions at small ratios but did not explain the data when the number of elements in one set became small. Small distributions are smoothed out entirely. This effectively challenges the assumption that an idealised orientation distribution is available. Simulations of Experiment 3.6 demonstrate that a ZC model + centroid model can deal with real orientation distributions.

Examining a typical stimulus for higher ratios as shown in Figure 3.6d-e, it is apparent that, as the numerosity ratio of the two distributions increases, the judgement from Experiment 3.2 becomes closer to a "search" task as described in the literature on visual attention (e.g. Treisman and Gormican, 1988). In Experiment 3.2 it appears that, in the case of multiple targets, threshold orientation difference between target and distractor may be well modelled by predictions based on change in the orientation variance of the texture. It is interesting to note that, for example, a variance model would predict a non-monotonic dependence of performance upon distractor number (shown by, for example, Sagi and Julesz, 1987, Sagi, 1990). From statistical considerations a small total number of texels leads to a poor estimate of variance, producing poor performance which improves with increasing texel number. However, as number increases, the influence of a single distractor on the global variance estimate will decrease. Subjects performance may be determined by the total of these two sources of error. The suggestion is that the underlying statistics of texel orientation may explain some visual search effects without recourse to specific, spatial interactions.

This chapter has assumed an orientation representation which ideally reflected the orientation probability distribution which generated it. How this might be achieved has not been considered. In particular in using texture fields of constant size the question of the spatial dimensions of any "region of integration" has not been addressed. Integration may occur upon local orientation measures, e.g. by spatial summation of similarly oriented receptive field outputs. Alternatively, integration may occur implicitly through the use of visual features grouped through the application of successively finer Laplacian-of-Gaussian filters (e.g. Watt, 1988). In both of these cases it seems likely that some measure of

reliability is employed to weight the contribution of extracted features. In the former case the ratio of outputs of differently oriented receptive fields at a point may be used. In the latter a mean-square deviation measure of a best fitting principal axes through the feature may be appropriate (Watt, 1991b).

The assumption of the availability of an idealised representation of attribute probability did not affect how well the candidate features accounted for the data. Given that estimates of the orientation bandwidths of linear filters in human vision are in the order of $10 - 20°$, this suggests that smoothing of orientation measures may occur to increase accuracy. This has also been suggested for a neural network implementation of the shape from texture model proposed by Blake and Marinos (1990).

In conclusion, this chapter strongly constrains a representation of texture orientation information. The variance and centroid of an orientation distribution are available to subjects. Multiple sets can be perceived and independent information about them is represented.

4 | MEAN ORIENTATION JUDGEMENTS USING LINE AND DOT TEXTURES

4.1 OVERVIEW

Chapter 3 considered how local orientation estimates are combined to form a representation suitable for subsequent texture processing. It was assumed that accurate measures of the local orientation of the elements comprising a texture are freely available. This chapter and the following three, consider exactly how texture elements, reflecting local orientation structure, are extracted from natural and artificial textures. It is proposed that oriented filtering in combination with the image description scheme of Watt (1991b) is an effective computational scheme for deriving texels. Such a model is also shown to provide a reasonable account of human performance on tasks requiring texel extraction.

Section 4.2 briefly considers some proposed schemes for deriving texels. A critical, and largely unaddressed, problem for most of these models is how to select the scale of analysis. Here the scale selection problem is treated as an instance of the generic *grouping* problem in vision and psychophysical tasks are employed which manipulate the difficulty of texel extraction by varying the "strength" of local grouping. A class of oriented random dot textures, known as *Glass patterns* (Glass, 1969), are used. These textures are useful because texels reflecting the orientation of these patterns emerge at a particularly narrow range of spatial scales. Unlike spatially band-pass textures, other scales contain information which, if relied upon, would actively disrupt an estimate of orientation. Because of this, Glass patterns are used to study the degree to which observers (and models) integrate across scale to form orientation estimates.

Psychophysical experiments are reported which examine the effects of element length, element density, and orientation variance on the judgement of the orientation of translational Glass patterns and line textures. In Chapter 5 data from these experiments are compared to the performance of a number of schemes for deriving orientation estimates.

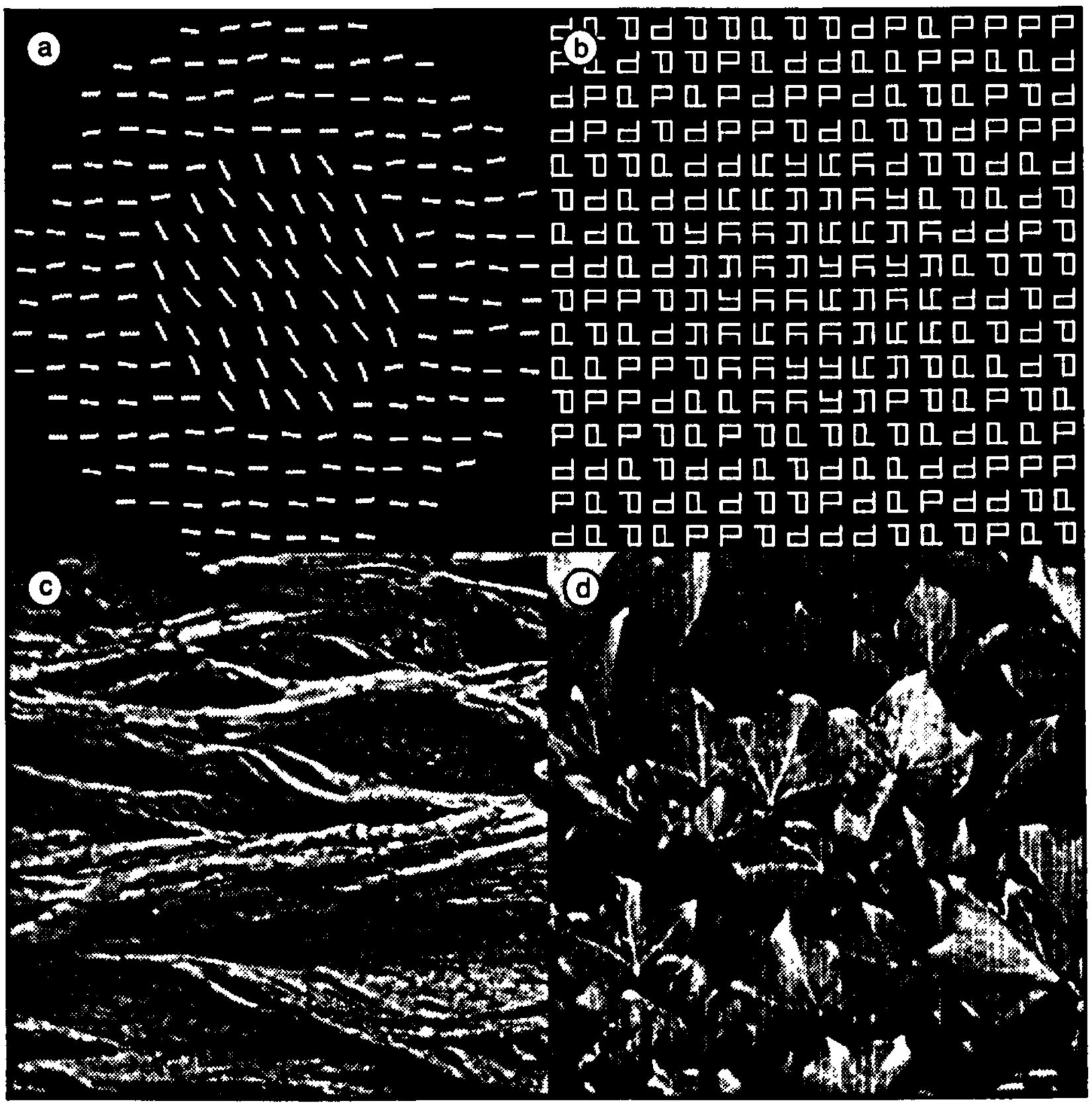


Figure 4.1. (a-b) Typical stimuli used in the psychophysical study of boundary detection. Texture elements are easily isolated. This is not always the case for natural textures, e.g. (c-d).

4.2 Finding texture elements

Many theories of texture segmentation (e.g. Julesz, 1981; Marr, 1982; Beck *et al.*, 1983) use statistical differences between the attributes of local texture elements. Texture boundaries exist where significant differences arise in texel density, size, etc. Theories of the derivation of shape from texture use the variation in the orientation (Witkin, 1981; Blake and Marinos, 1990), or density (Gibson, 1979) of texels to signal local surface orientation. Thus texels must also be *oriented* (Blake *et al.*, 1993). Both of these approaches assume that texels can be isolated automatically from a scene by human vision.

Reflecting the view that the extraction of texels is not an interesting component of texture processing, the artificial textures used in psychophysics have attempted to make the isolation of texels trivial for subjects. Figures 4.1a-b show that this is achieved by constructing textures from spatially distinct (located on a perturbed grid) micro-

patterns which are much brighter than the background. Such patterns could be isolated by a scheme as simple as grey-level thresholding followed by extraction of connected regions. It is implicit in all work on texture segmentation that experimental stimuli, such as Figures 4.1a-b are, to some degree, an approximation to texture boundaries in natural images. Otherwise, why is the segmentation of these images interesting? However, when one considers images of natural textures, like Figures 4.1c-d, it is apparent that there are no obvious texture elements. Even though it is acknowledged that "there is no known algorithm that can successfully detect texels from a natural image" (Aloimonos, 1988), the development of algorithms for deriving information from texture continues, with each theory making different assumptions about what information texels can give. Given that the form of the texels which can be derived automatically from natural images will considerably constrain further processing by, for example, shape from texture algorithms, it is strange that the problem of texel isolation has remained largely unconsidered.

Recently, however, Laplacian-of-Gaussian ($\nabla^2 G$) filtering has been proposed as a computationally efficient method for extracting texture elements from images (Voorhees and Poggio, 1987; Wen and Fryer, 1991; Vilnrotter *et al.*, 1986): being an isotropic filter, only one convolution is required at each spatial scale. The elements are typically derived from zero-crossings (e.g. Witkin, 1981), half-wave rectification (Voorhees and Poggio, 1987) or by fitting templates to the output (Blostein and Ahuja, 1993). The relative merits of edge and blob based representations have been discussed elsewhere (Blostein and Ahuja, 1993). Given the output of these filters followed by, e.g. thresholding (Figures 4.2b-d), it is possible to construct a *symbolic* description of each blob in terms of parameters such as length, mass, etc. (e.g. Voorhees and Poggio, 1987; Watt, 1991b) giving a "primal sketch" type of representation (Marr, 1976). However, the adequacy of such a description for performing tasks which require texture information has only been shown using demonstrations. What is not known is whether these systems are similar to the way that human vision isolates texels.

A second unaddressed question relates to a parameter which is usually ignored in theories of texture perception based on spatial filtering: the filter scale itself. Consider Figure 4.2 which shows an image of tree bark and three $\nabla^2 G$ filtered and thresholded versions of (extending over 3 octaves of scale). Each of these images highlights orientation structure in the texture at a different spatial scale. The output of the fine-scale filter, Figure 4.2b, captures detail of the shallow ridges in the top-left hand corner of the image,

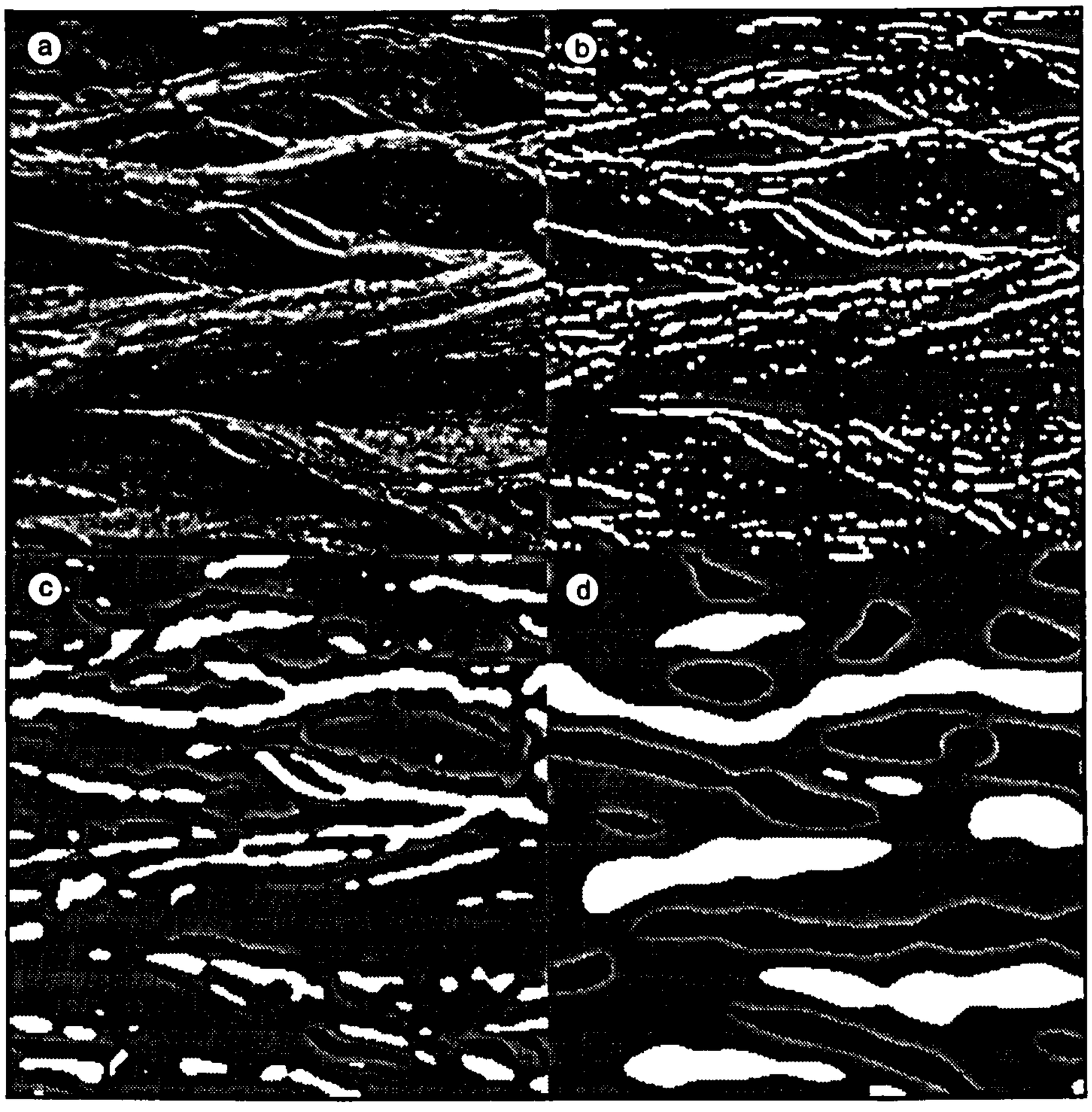


Figure 4.2. (a) Bark texture (b-d) Output of $\nabla^2 G$ filters with standard deviation (s.d.) of (b) 1.41, (c) 4.00 and (d) 11.31 pixels. Filtered images have been thresholded at one grey level standard deviation and all grey levels above and below the mean grey level replaced with white and black pixels respectively.

for example. The middle spatial scale, Figure 4.2c, is less sensitive to fine surface creases but begins to highlight the overall direction of flow in the image, from left to right. The coarsest scale obliterates all fine detail but captures the flow due to the very deep creases in the bark. When presented with this image, what determines which scale is used to derive texels? As discussed in the introductory chapter, it is in the nature of texture that the answer to this question depends on what the visual system is to do with the texels.

In this chapter the problem of deriving texture elements from a class of patterns known as *texture flows* (Zucker, 1982) is considered. A texture flow, such as a patch of fur or wood-grain, is rich in orientation structure, but is not necessarily defined by clear, extended contours. The perception of structure in a fur pattern, for example, involves perceptual "filling-in" of some sort. Individual hairs pass in and out of occlusion yet the percept of orientation structure is dense and complete (Zucker, 1982). These patterns are interesting, not only because orientation is an essential cue for many visual tasks, but also because they are so completely defined by their orientation structure that they allow inferences to be made about the operation of visual mechanisms upon them.

Assume that whatever process is responsible for estimating local orientation attempts to maximise the *accuracy* of its estimate. Statistical wisdom indicates that the source of information being used should have minimal variance. This suggests a simple strategy for determining any free parameters (such as filter size) of a texture processing model automatically: *set the parameters so that the derived texture elements have minimal orientation variance*. By examining a symbolic image description, from a system such as that of Watt (1991b), the local orientation variance of a given set of blobs may be calculated, and so an estimate of the reliability of a particular spatial scale, for estimating local orientation, may be assessed. By using texture flow patterns, rather than, for example, embedded patches of texture, *the required output* of a texture processing task is constrained sufficiently that the task constraints may be incorporated directly into the model. The idea that a model based on this assumption reflects *human* visual processing is tested in this chapter. In order to describe psychophysical data on the perception of texture flows, an artificial texture flow must now be introduced: the Glass pattern.

4.3 Visual grouping and Glass patterns

Much of the psychophysical work on texture flow has employed random dot patterns. These are used because they are simple to generate, and because dots have generally been

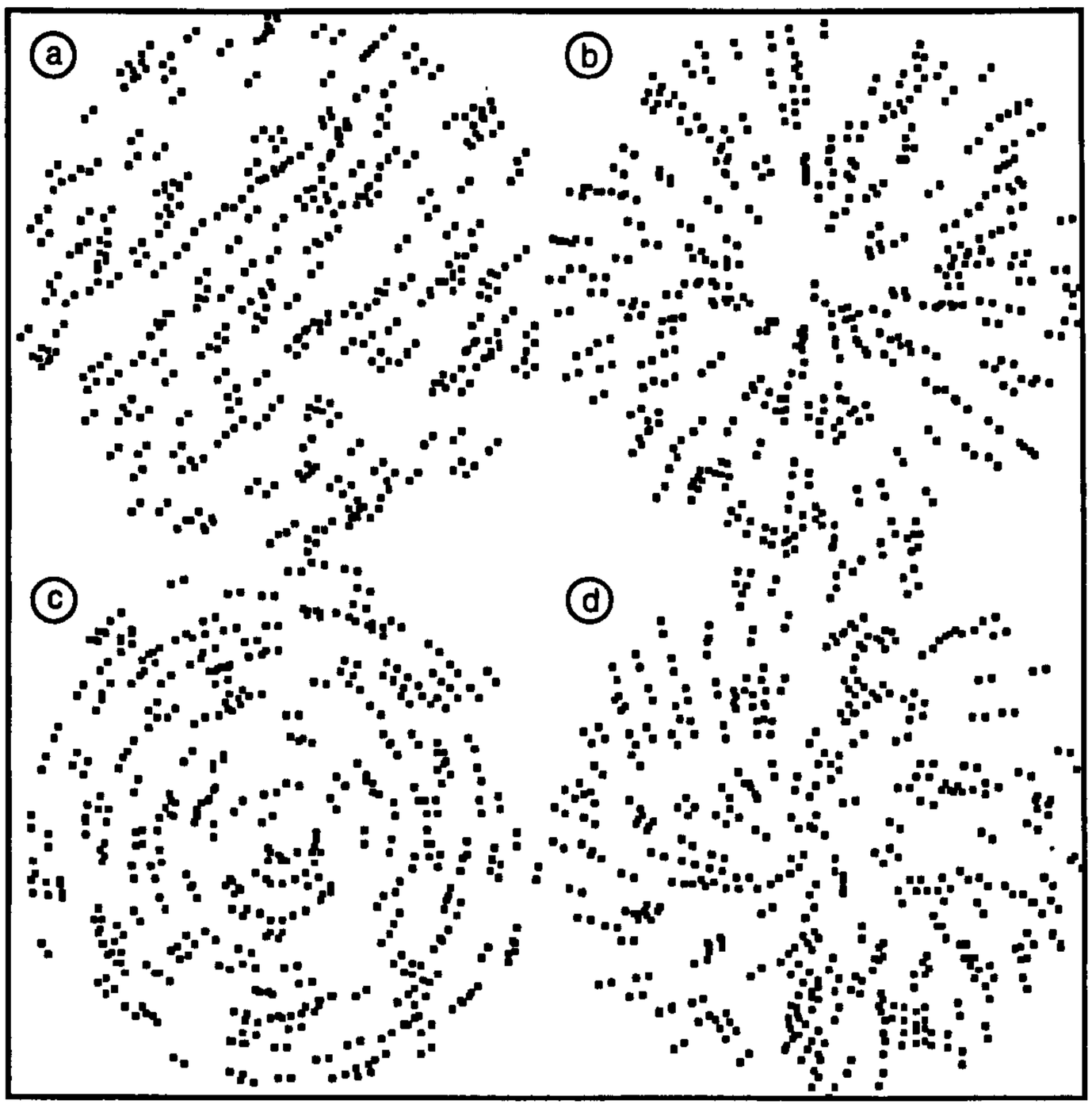


Figure 4.3. Glass patterns composed of two superimposed sets of dots, the second set being the first set transformed by (a) a translation, (b) a dilation, (c) a rotation and (d) a spiral.

treated as though they are approximations to the mathematical notion of a point, i.e. objects without dimensions, defined purely by *location*. Figure 4.3 shows some examples of *Glass patterns* (Glass, 1969). These patterns have properties which prove to be very useful for studying texel extraction in the context of spatial scale.

Glass patterns are composed of the superimposition of one or more copies of a field of randomly distributed features (e.g. dots) onto the original, where the copy is a geometric transformation of the original (Glass, 1969; Glass and Perez, 1973). The visual impression is of compelling oriented structure with dot pairs (*dipoles*) aligned along the direction of the local transformation. Glass patterns are interesting stimuli for a number of reasons. Firstly, they approximately isolate the selection of orientation from other visual processes due to, e.g. contrast and size differences (Zucker, 1982). Secondly, these patterns only contain the perceived structure at a narrow band of spatial scales, which makes them ideal for investigating scale selection processes. Thirdly, there have been a large number of computational and psychophysical investigations of the perception of structure in these patterns. Finally, although apparently simple textures, they remain a largely unexplained phenomenon; a number of models have been proposed, none of which have been shown to be completely satisfactory.

4.3.1 Psychophysical work

The work done with Glass patterns by Kent Stevens (1978) is important not only as an investigation of visual grouping, but also as one of the first papers to combine psychophysical experiments with computational modelling of visual processes. It is also one of the few explicit investigations of Marr's (1976; 1982) theory of the primal sketch, and was used as practically the sole evidence for grouping processes in Marr's full exposition of his theory (Marr, 1982).

Stevens (1978) used patterns composed of dots whose locations formed a perturbed grid. Dipoles were not permitted to fall in such a way that alignment with dipoles from neighbouring cells could cause "chains" of more than two elements to occur. A rating judgement of "pairedness" was used to determine the maximum dipole separation for which structure was rated to be present. Dot density varied from 0.5 to 44 points/degree^2 (from 65 to 580 total dots) and the results indicated that, regardless of pattern type, if more than 2 or 3 points lay closer to a dot than its corresponding dot, then structure was

not rated as present.

Jenkins (1983), examining subjects' discrimination of Glass patterns from unorganised noise, produced results directly contradictory to Stevens' experiment, showing that there is no effect of altering dot density over a range of densities from 6.5 to 26.0 points/degree^2 (corresponding to approximately 64, 129 and 258 total dots). Jenkins (1983) accounts for Steven's low estimate of the amount of tolerable noise by assuming that Stevens' subjects were conservative in their subjective rating of "pairedness". Another explanation for this inconsistency of findings is that the stimuli used in Stevens (1978) could not contain accidentally aligned dipoles, or "chains',' of two or more dot pairs. The presence of these (essentially low spatial frequency) features could be an indication of structure when dipole separation becomes large. Jenkins (1983) goes on to claim that stimulus field diameter is the critical factor in perceiving these patterns, independent of viewing distance.

Maloney *et al.* (1987) also measured the discriminability of Glass patterns from patterns of randomly oriented dipoles, but they varied the number of unpaired noise dots added to the original pattern. Using two dot separations and a number of dot densities, they showed a Weber law relationship between the number of dipoles in the pattern and the maximum number of unstructured dots which could be tolerated at a particular level of performance. Their data show that detection at a level of $d' = 1.0$ is possible with more than six noise dots closer to a dot than its partner. They suggest that this result argues against the neighbourhood approach of Stevens (1978) but, again, it may simply indicate that the stimuli and task used in Stevens (1978) preclude comparison of data with any other work. Maloney *et al.* (1987) also state that variation of up to $\pm 12°$ in dipole orientation has no effect on performance.

4.3.2 Stevens' model

Stevens (1978) proposed a representation of structure in Glass patterns. Since correctly paired dots will form dipoles which are locally parallel, Stevens proposed a model which calculates all possible pairings of a dot in a small surrounding region, and all possible pairings of those dots in a similar sized region, and uses the most frequently occurring orientation to select the correct correspondent. The operation of the model is illustrated in Figure 4.4.

There are a number of problems with this model. Firstly it will be shown (in Chapter 5) that Stevens' estimate of tolerable noise, which is a critical part of the rationale behind his

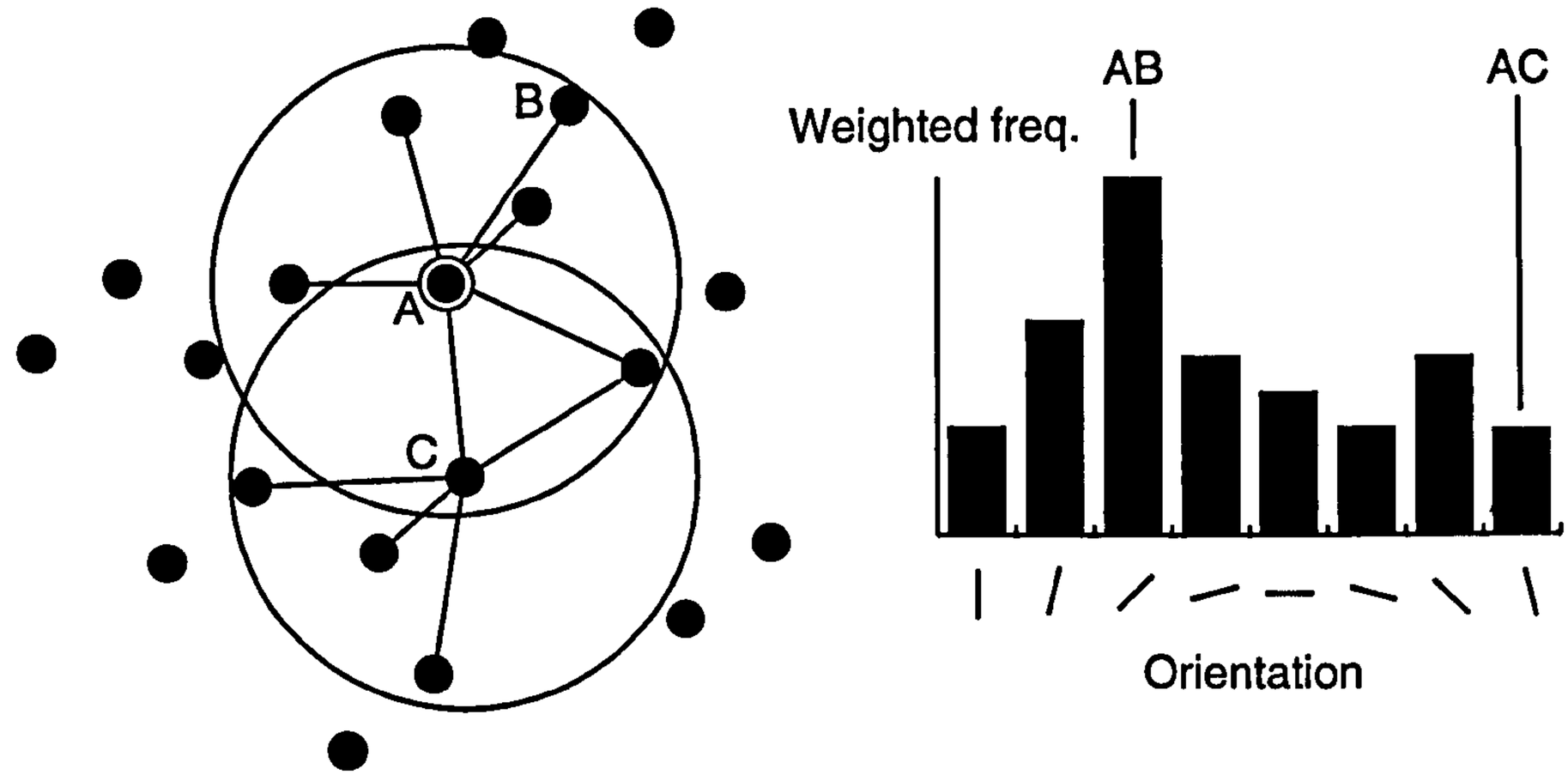


Figure 4.4. Stevens' algorithm for virtual line construction in Glass patterns. A dot A is connected to all other dots in a neighbourhood around it, and these dots (e.g. C) are connected to all dots lying in a neighbourhood centred on their locations. The orientation of all matches are histogramed and the peak orientation used to select the actual match to A (in this case B).

model, is very low if one allows chains in the Glass patterns. Patterns containing accidental alignments of dipoles signal structure not only through the orientation of dipoles, but through larger clusters of dipoles in the image. How can a model based on local parallelism account for this performance? It will further be shown that the good performance of the model on Glass patterns is largely due to the fact that the dot locations of elements of the test patterns fall on a perturbed grid. When dots are allowed to fall anywhere in the pattern, the model tends to perform poorly. Finally, the model is only applicable to dot patterns where the location of each element is clearly defined. How does one derive features from natural images?

4.3.3 Spatial filtering models

As mentioned in Section 4.2, grouping in textures can be achieved through the application of spatial filters (e.g. Zucker, 1982; Kass and Witkin, 1985). These approaches work because oriented receptive fields are not edge detectors, but are merely sensitive to contrast differences along their axes. Zucker's (1982) model, for example, estimates the vector field of tangents in an image via oriented filtering and a cooperative interpretation algorithm. Integral curves are then found throughout the field. The initial filtering is motivated by simple cell receptive fields in V1, and occurs at a variety of scales. Dominant orientation

at a point is calculated by selecting the orientation of the filter giving maximal response.

Kass and Witkin (1985) isotropically smooth the input image and calculate derivatives in the x and y directions, to give C_x and C_y, respectively. Given a further (Gaussian) weighting function, W, the local orientation (ϕ) is calculated as:

$$\phi = \frac{1}{2}\tan^{-1}\left(\frac{W * 2C_xC_y}{W * (C_x^2 - C_y^2)}\right)$$

This approach makes two key assumptions. Firstly, that the output of the filtering stage will always be zero-mean, which is only true as long as the initial filtering stage does not operate at too coarse a resolution. Secondly that calculating the mean of the variance of the filter output, as a function of orientation, codes local mean orientation. The argument for this is informal.

It is apparent that models for deriving orientation structure differ principally in the way in which the orientations of elements are obtained. Kass and Witkin (1985) use the image gradient of the rotated and Gaussian smoothed image. Stevens (1978) and Caelli *et al.* (1978) use virtual lines and Zucker (1982) uses oriented Difference-of-Gaussian operators.

Prazdny (1986a) pointed out two principal problems with filtering models. The first is a specific problem they have with *opposite contrast* dot patterns which will be considered in some depth in Chapter 7. The second is one of the questions being addressed in this chapter: how is filter size selected?

The selection of scale is not a trivial part of the grouping process in the case of Glass patterns because these patterns only contain useful information at a narrow range of spatial scales. Figure 4.5 shows scale-orientation histograms of a Glass pattern and a line texture. These histograms use blob descriptions (Watt, 1991b), derived with a range of filter sizes, to show how orientation and length statistics change with the spatial scale of analysis. The vertical band running from the middle to the bottom of the line texture orientation histogram (middle row, left) indicates that a wide range of fine spatial scales would give an accurate estimate of the mean orientation. The orientation histogram of the Glass pattern (middle row, right) has no such line. The "hour-glass" shape of this histogram shows that there is only a narrow range of filter sizes which will produce the correct interpretation of the pattern. Notice that the length-scale histogram of the Glass pattern shows that mean blob length increases with scale. This information does not appear to be of much use for the selection of scale. Indeed the primacy of orientation

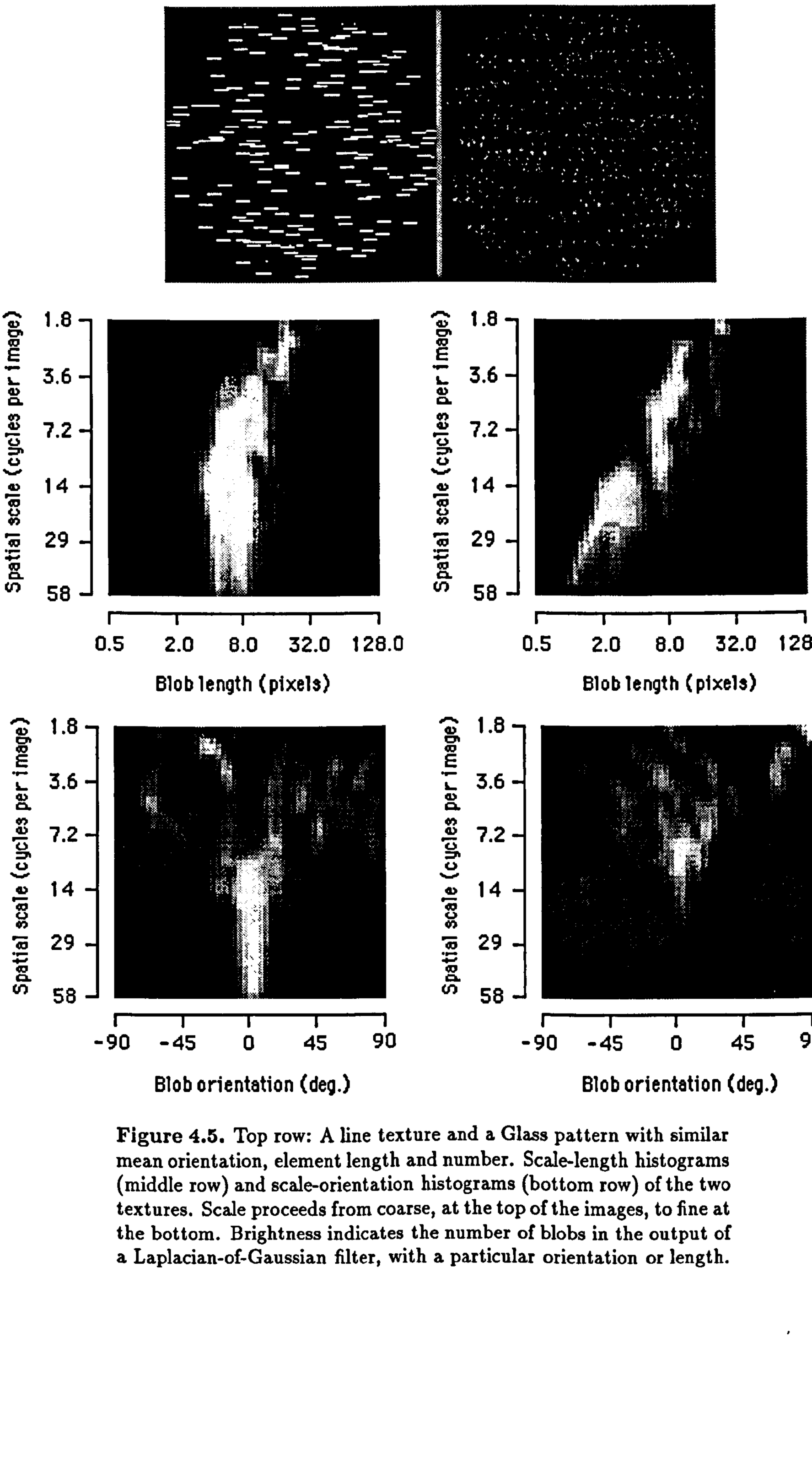


Figure 4.5. Top row: A line texture and a Glass pattern with similar mean orientation, element length and number. Scale-length histograms (middle row) and scale-orientation histograms (bottom row) of the two textures. Scale proceeds from coarse, at the top of the images, to fine at the bottom. Brightness indicates the number of blobs in the output of a Laplacian-of-Gaussian filter, with a particular orientation or length.

information in the processing of these patterns has been demonstrated psychophysically by Caelli and Julesz (1979), who showed that orientation, but not length, variance determined discrimination strength of patches of dipoles.

Prazdny (1986a) points out that to overcome this problem there must be "an evaluating agent" looking at the output of the filters at various scales which, he suggests, is like the Gestalt notion of "Prägnanz". All of the models described so far either make assumptions about setting of neighbourhood size (Stevens, 1978) or filter size (Kass and Witkin, 1985; Zucker, 1982), or cannot explain global organisation effects at all (e.g. autocorrelation models such as Maloney *et al.*, 1987).

Models have not considered the scale selection problem because processing has not been considered in the context of particular visual tasks. No one criterion for the selection of scale will suffice for all tasks. However, tasks may be considered which demonstrate a number of generally applicable criteria used by the visual system. In the next section a mean orientation judgement is used to define clearly such a criterion on statistical grounds: in this case the minimisation of local orientation variance.

The remainder of this chapter presents three experiments examining the extraction of the mean orientation of Glass patterns as a function of a number of stimulus parameters. In order to isolate the effect of grouping on the task, performance with textures composed of lines is also measured. Differences between line and dipole textures should be attributable to grouping uncertainty. Finally three models are used to simulate performance on these tasks: a model using Laplacian-of-Gaussian filtering, one using Difference-of-Gaussian, filtering and the Stevens' algorithm (Stevens, 1978).

4.4 General Methods

The following three experiments used the same basic method but manipulated different aspects of oriented textures. The independent variables examined were the length of elements (Experiment 4.1), the orientation variance of elements (Experiment 4.2) and the density of elements (Experiment 4.3). Experiment 4.1 and 4.2 also compare data from equivalent tasks using line textures.

In all the experiments a judgement of the mean orientation of oriented textures was used (the usefulness of this task has been discussed in Chapter 3). In this context, a mean orientation judgement was used with the goal of explaining performance, using a computational model, clearly in sight. Such a task places clear statistical constraints on

the output of any visual process for extracting local orientation structure. It is further hypothesised that signal detection or "structure-versus-noise" tasks (e.g. Jenkins, 1983; Wagemans *et al.*, 1993) may fail to discriminate between the performance of different models because of ceiling effects, i.e. there may be too many sources of information for performing such a task to separate predictions from different models. By considering the accuracy with which a subject can make an estimate of a single stimulus attribute, such as overall orientation, it is hoped that predictions from various models can be separated.

4.4.1 Subjects

Four subjects, aged 24-32 years, served as subjects in the experiments. All were experienced in psychophysical procedure. The author (SCD) and subject BJC were aware of the purpose of the experiments while the other two subjects (FJM and RAO) were naive. All subjects had normal or corrected-to-normal vision.

4.4.2 Procedure

The generation and presentation of stimuli, and the recording of subject responses were carried out on a Macintosh IIfx microcomputer. The display was a Formac ProNitron 80.21 colour monitor with a frame refresh rate of 75 Hz. The screen was viewed binocularly with natural pupils at a distance of two metres. All stimuli were presented in the centre of the screen, which was indicated by a pre-stimulus marker, for 100 ms, to prevent saccading during presentation. An ISI of 750 ms followed each response.

The subjects' task was a single interval, two alternative forced choice, and was to report whether the texture presented had a global orientation clockwise or anti-clockwise of a reference orientation. The reference orientation was vertical and was not presented to the subject, i.e. an internal reference was used.

Subjects responded by depressing one of two keys on the computer keyboard and could depress another key to get a second look at a stimuli. They were encouraged to use this facility only if they actually did not see the stimuli, e.g. by blinking.

An adaptive method of constant stimuli, APE (Watt and Andrews, 1981), was used to sample a range of mean orientations around vertical. Three runs of 64 trials were undertaken for each data point presented. Conditions were not interleaved. At the end of a block, probit analysis was used to estimate the standard deviation of the psychometric function for each run. The data points plotted are the arithmetic mean of these values,

and the error bars show ± 1 standard error.

4.4.3 Stimuli

All stimuli used in the experiments were approximately circular texture fields, with radii subtending 1.23 deg. (128 pixels), of either lines or dot pairs (dipoles). These fields appeared within a 2.46 deg. (256 pixel) square image. Texels appeared white on a black background, and were distributed randomly throughout the field.

Line elements were anti-aliased, using 16 grey levels. Component dots of the dipoles were individual pixels which subtended approximately 35 arc seconds. No anti-aliasing of dipoles was used. If the orientation of a dipole required that one of the component dots be placed in a position between the discrete pixel locations available, the nearest pixel location was used. As a consequence, in the experiments employing a constant dipole length (Experiments 4.2-4.3), an apparently large value of 8 arc min. (approximately 14 pixels) was used (see Figure 4.8 for examples). At this length, cues as small as 4° can reliably be presented. Other stimulus parameters were selected to avoid any ceiling effects that might be associated with such a minimum cue size.

The orientation of elements of both line and dipole textures were drawn from Gaussian random distributions. Apart from Experiment 4.2, where the effect of orientation variance was investigated, distributions had a standard deviation of 8.0°. This value was used because pilot studies indicated that such a level of variability, with textures composed of 8 arc min. long dipoles, elevated performance by around 50% for two subjects, bringing their thresholds closer to the minimum presentable cue.

Vertical patterns were chosen to avoid problems due to the well known oblique effect (Appelle, 1972). Vertical patterns were chosen over horizontal because of the established advantage for vertical Glass patterns over horizontal (Jenkins, 1985).

Textures in Experiments 4.1 and 4.3 were composed of 512 lines or dipoles, corresponding to an average density of 89.4 elements/degree^2 (Experiment 4.3 explicitly investigated the effect of element density). A pilot study indicated only a small effect of element density on thresholds for the line and dipole textures. The number of elements selected, 512, was chosen because theories based on symbolic matching of tokens (e.g. Stevens, 1978) predict that performance should be low at densities where each dot will have an average of 6 dots closer to it than its correspondent (Stevens (1978) states that if more than 2 or 3 dots lie closer to a given dot than its correspondent, then local structure cannot be

perceived).

4.5 Experiment 4.1: Effect of element length on the judgement of mean orientation

The first experiment investigated the effect of element length on the judgement of the mean orientation of translational line textures and Glass patterns. The experiment was performed primarily as a source of data to compare a number of models of local orientation calculation. By varying element length one can look at the performance of various matching mechanisms in the presence of noise due to the proximity of other, uncorrelated dots. Furthermore subjects thresholds for line textures will give an estimate of the absolute limits on the estimation of the mean of a set of oriented elements in the absence of matching uncertainty.

Previous work examining the effect of dipole length on the perception of Glass patterns has used rating judgements (Caelli, 1981) or discrimination of structure from noise tasks (Jenkins, 1983; Wagemans *et al.*, 1993). Caelli (1981) asked subjects to rate the "perceived extent" of Glass patterns, i.e. how far from the centre of a rotational field structure was still visible, as a function of the range of rotation of the patterns. He found that performance falls steadily as a function of increasing angle of rotation. However, the unreliability of subjective ratings preclude comparison with quantitative measures of performance.

Jenkins (1983) argued that perception of structure in Glass patterns fell into three categories according to dipole length. At small separations individual dipoles are visible, leading to a compelling perception of flow. At larger separations patterns become "striated": individual dipoles are not visible but the pattern still has overall orientation. Finally, when dipole length exceeds some critical value, no structure is perceived. Jenkins (1983) used patterns composed of approximately 258 dipoles and determined that discrimination from noise falls as dipole length is increased. Jenkins (1983) quotes a limit of 1.4° for 50% correct discrimination of signal from noise.

Wagemans *et al.* (1993) also measured the discriminability, d', of Glass patterns from noise patterns as a function of dot separation. Patterns were of extremely low density (16 dipoles) and were viewed for 100 ms. In order to test the hypothesis that "higher-order structure", in the form of quadrilateral groups of dots, is important to the perception of structure, the effect of adding variation to dipole length was also examined. It was found that increasing dipole separation produces worse discrimination. Length variation was

also found to have a detrimental effect on signal detection.

Wagemans *et al.* (1993) and Jenkins (1983) have both interpreted their data as evidence for the different models proposed in each. The finding of Jenkins (1983), that there is a critical dipole length for perception of structure, seems to support a matching mechanism using spatial correlation; although a variety of other models (e.g. Stevens, 1978) would also predict such a limit. In Wagemans *et al.* (1993), the efficiency of subjects at performing the discrimination task (d' as high as 4.6, i.e. 100% correct discrimination) is in no way matched by the simulated annealing model proposed (maximum of around 85% correct discrimination). Although the model produces "the same rank ordering of performance levels" (Wagemans *et al.*, 1993) this is an extremely weak criterion for assessing the validity of a model. This is especially so, given that the model has free variables, such as the grouping neighbourhood size, which can be manipulated to match human data. It may be concluded that quantitative explanations of the effect of dipole length on perception of structure in Glass patterns have been wholely inadequate.

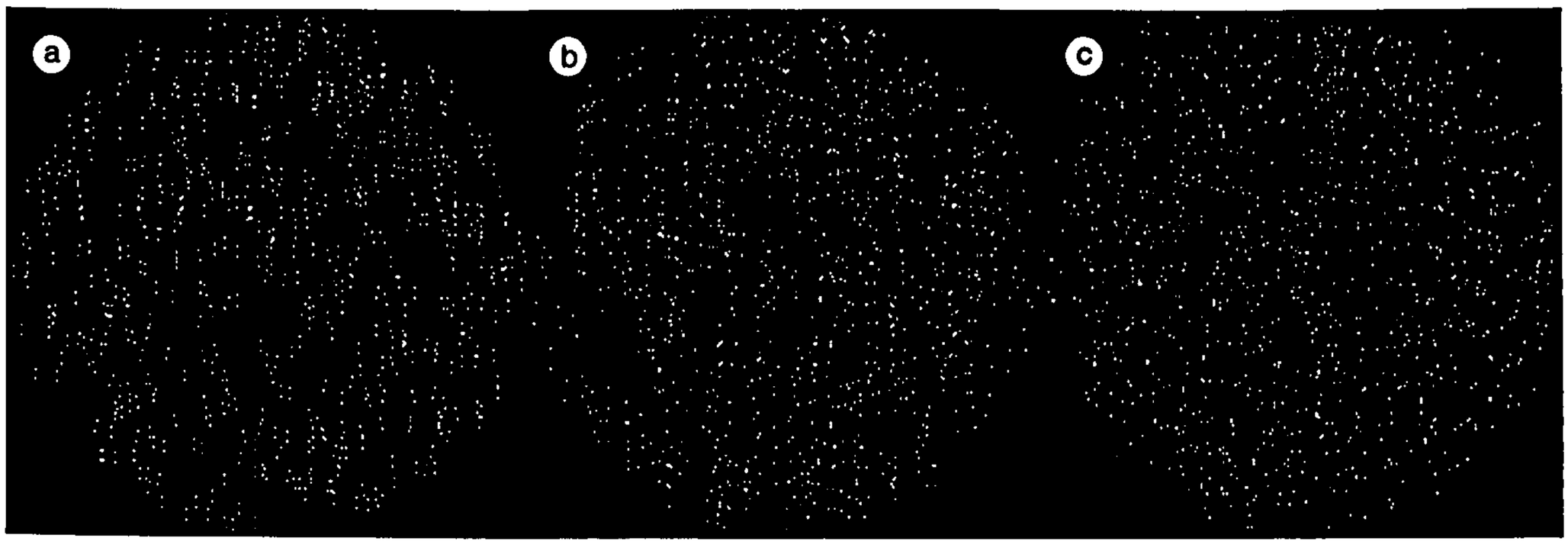


Figure 4.6. Examples of the stimuli used in Experiment 4.1. The patterns shown have dipole lengths of (a) 2.8, (b) 5.6 and (c) 11.3 arc min.

4.5.1 Stimuli

Dipole and line textures were used, all composed of 512 elements each 8 arc min. long. Element orientations were drawn from a Gaussian random distribution with standard deviation 8.0°. A range of dipole lengths was tested from 1.41 arc min. to 32.0 arc min. in multiplicative steps of $\sqrt{2}$. Examples of the stimuli are shown in Figure 4.6. Stimuli were presented for 100 ms.

4.5.2 Results

Threshold offsets for the mean orientation judgement as a function of element length are shown for two subjects in Figure 4.7. The accuracy of judging the mean orientation of line and dipole textures improves rapidly with element length up to 4-5 arc min. This improvement is probably due to two factors. Firstly the dependence of the accuracy of orientation estimates on the aspect ratio of the figure (Westheimer, 1981; Vassilev *et al.*, 1981). Secondly, at small dipole separations the cue probably has to be relatively large to overcome the problem, at small cue orientations, with the discrete location of the pixels comprising each dipole. The second factor does not appear to be a major contributor to poor performance at small element lengths because performance using the line textures, which are anti-aliased and so do not suffer from this problem, closely follow that with the dot textures.

As element length increases above 4-5 arc min., accuracy for judgement of line texture orientation improves until it asymptotes at around 2.0°. Judgement of the orientation of Glass patterns, however, quickly breaks down as dipole length increases beyond 8 arc min. This breakdown in performance is due to uncertainty in matching the dipole components. The task becomes impossible with dipole textures at around 23-32 arc min.

This deterioration in performance as dipole length increases, confirms the general finding of Jenkins (1983) and Wagemans *et al.* (1993)[1]. Experiment 4.1's estimate of the distance at which orientation estimates break down is 23-32 arc min., at a viewing distance of 2.0 metres. Jenkins (1983) estimate of the maximum dipole length facilitating a structure-versus-noise task is 1.4° measured at 57.3 cm. Scaling Jenkins' (1983) estimate (i.e. making the strong assumption that viewing distance will have little affect on performance) produces an estimate of 24.0 arc min.: closely in agreement with the result of this experiment.

4.6 Experiment 4.2: Effect of orientation variance on the judgement of mean orientation

The aim of this experiment was to determine the effect of adding local orientation variance on the judgement of the mean orientation of a Glass pattern. This form of noise is important because one would expect that a model relying on local orientation statistics

[1]Further comparison with the Wagemans *et al.* (1993) is not possible as the distances quoted in this paper are in display pixels and the size of the display is not given

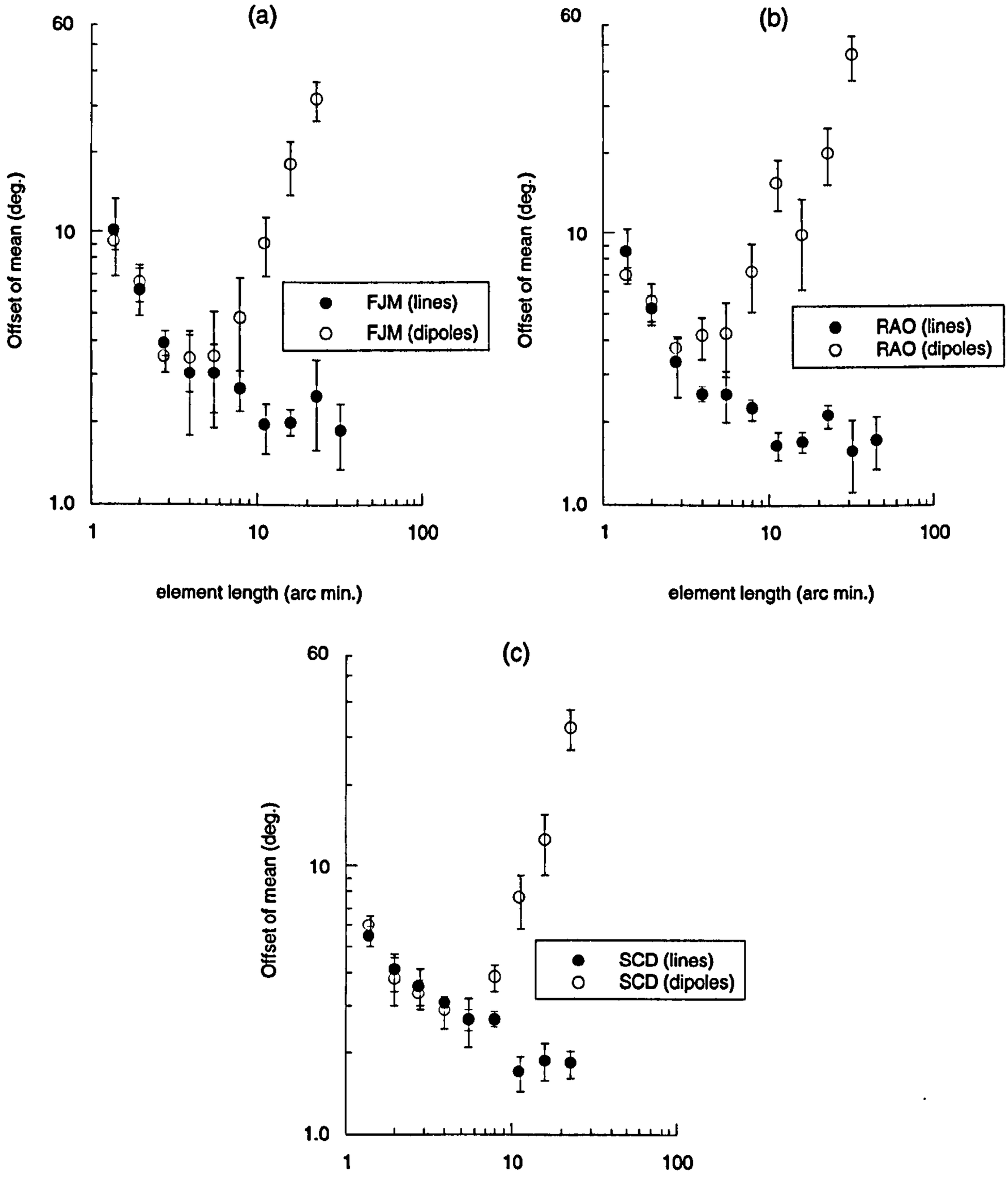


Figure 4.7. Threshold mean orientation offsets for judgement of the orientation of line and dipole textures, as a function of element length.

(e.g. oriented filtering) will be critically affected by changes in these statistics.

The effect of adding small random rotations to dipole orientations was first observed by Glass and Switkes (1976) who informally demonstrated that such noise degraded perception of structure. This, they claimed, was consistent with the physiological model described by Glass (1969): the range of dipole directions now exceeds the specificity of a single orientation column and the excitation required to perceive structure is not achieved. Maloney *et al.* (1987) indicated that this range must be less than $\pm 12°$ because such a range of orientations does not significantly affect perception of structure. There is no other psychophysical evidence, to my knowledge, concerning perception of structure in the presence of local orientation noise.

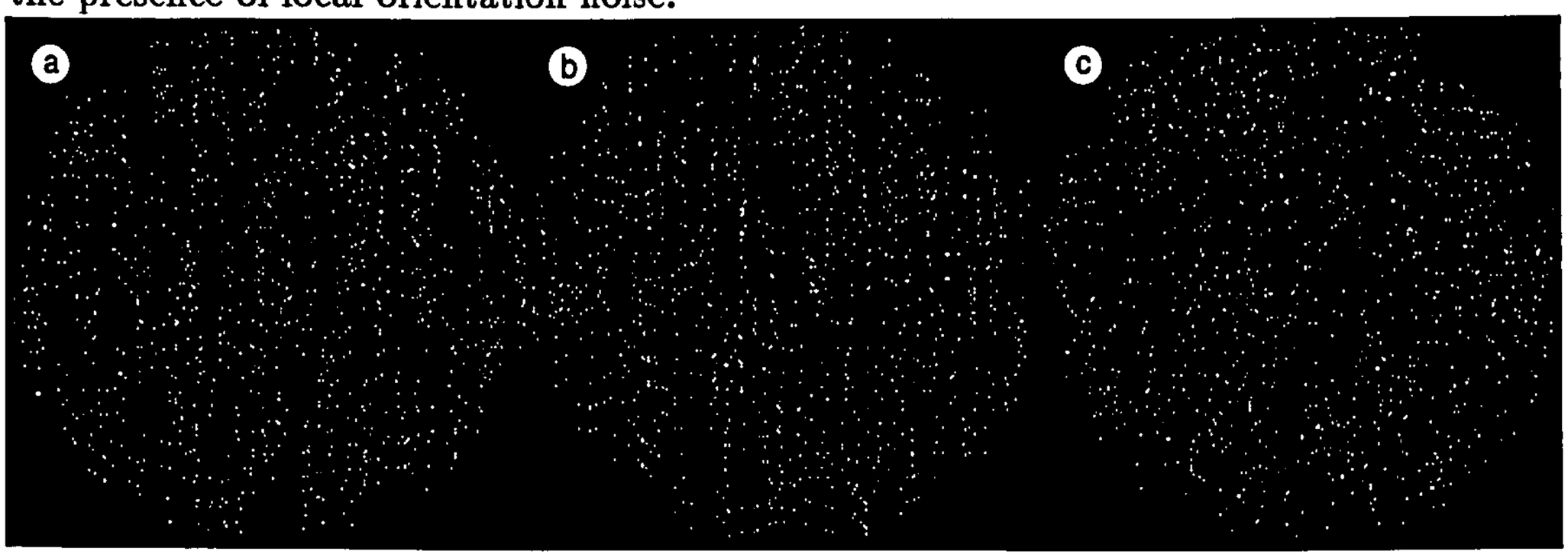


Figure 4.8. Examples of the stimuli used in Experiment 4.2. The patterns shown have orientation variability with s.d of (a) 1.0°, (b) 4.0° and (c) 16.0°.

4.6.1 Stimuli

Line and dipole textures, similar to those used in the previous experiment, were used, except that the length of elements was fixed at 8.0 arc min., and the local orientation variance of elements was systematically varied. Dipoles had Gaussian distributed orientations with a standard deviation of from 1.41° to 32.0°, sampled in multiplicative steps of $\sqrt{2°}$. Examples are shown in Figure 4.8.

4.6.2 Results

Threshold offsets for mean orientation judgement as a function of dipole orientation variance are shown for three subjects in Figure 4.9. Subjects typically achieve thresholds as low as 3.0° for the dipole textures and 2.0° for the line textures: an impressive level of

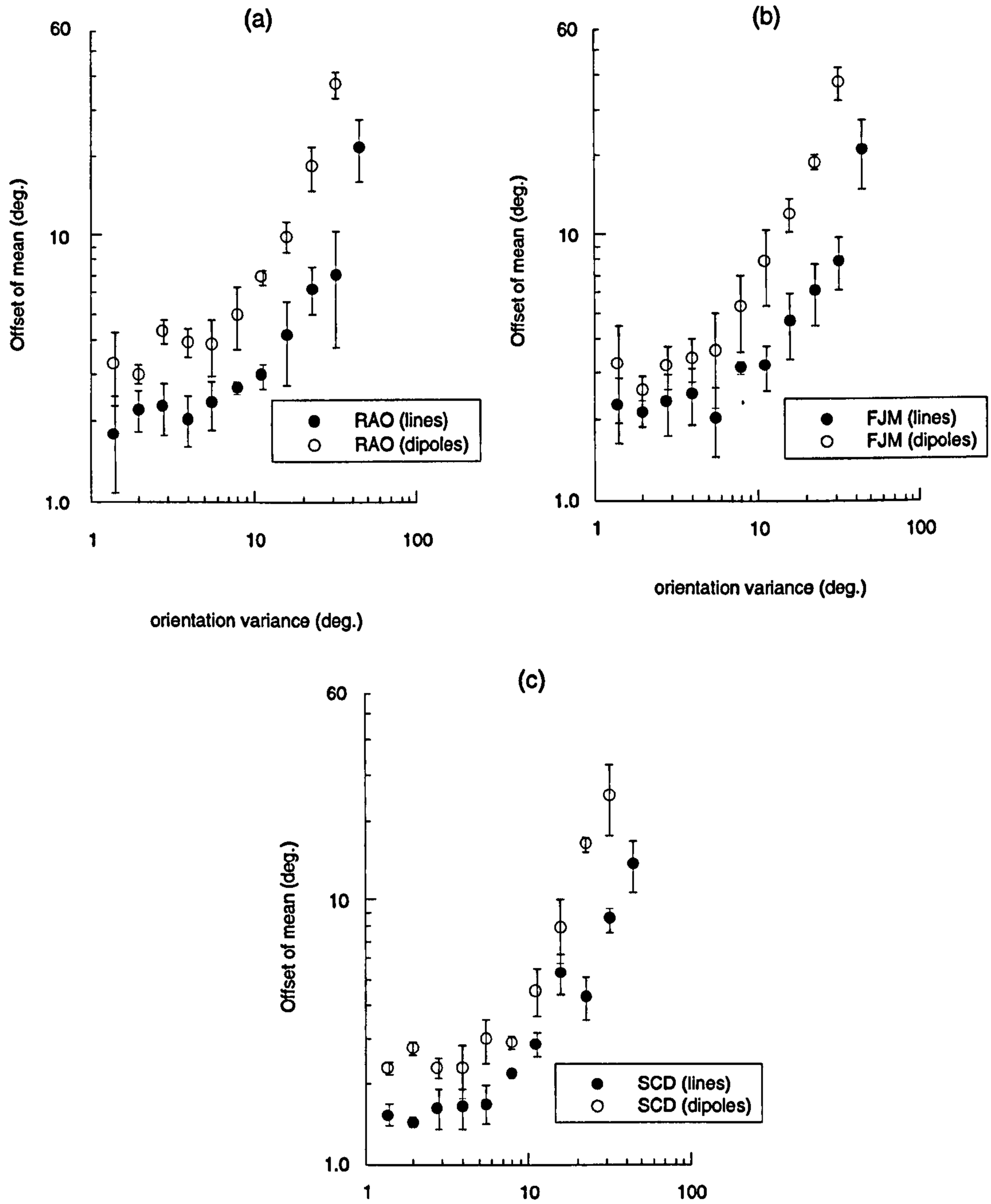


Figure 4.9. Threshold offset for mean orientation judgement of line and dipole textures, as a function of element orientation variability.

performance given the large separation of dots and high density of dots. The pattern of results for the line and dipole textures are similar except that there is a uniform shift of the functions from the lines to the dipoles. Such a uniform shift on log axes indicates a multiplicative effect of matching uncertainty in the Glass patterns.

It is clear that there is little effect of adding orientation jitter on either the line or dipole textures until the standard deviation exceeds about 8°. This figure is in accord with the value of ±11° quoted in Maloney *et al.* (1987), which is equivalent to Gaussian distributed orientations with a standard deviation of 7.8°. Such a figure seems to indicate that there is inherent noise on the system which limits the accuracy of estimating mean orientation at very low levels of jitter. It is quite possible that this noise is due to characteristics of the filters employed to extract structure. Beyond this level of orientation variance, performance deteriorates in an approximately power law relationship with orientation variance.

4.7 Experiment 4.3: Effect of element density on the judgement of mean orientation

For models based on "neighbourhood" matching, the number of dots lying closer to a dot than its correspondent is critical. However, Jenkins (1983) reported that signal density has no effect on the maximum dipole length for detecting structure in vertically translated Glass patterns. Element density has also been shown to have little effects on the detection of symmetry in dot patterns (Jenkins, 1985), and on the displacement limits for detection of motion in random binary luminance patterns (Morgan and Fahle, 1992).

The aim of this experiment was to determine the effect of the number of elements making up a Glass pattern on the accuracy of judging the mean orientation. Pilot trials indicated that density had little effect with either the line or dipole textures, so this experiment was only carried out using Glass patterns, where the effect of this variable will make clear predictions for models of texel extraction.

4.7.1 Procedure

A mean orientation judgement was again used in this experiment; identical methods were used to the two previous experiments. The only difference was that the number of dipoles in the texture was systematically varied. Note that since the size of the display was fixed, the density of the pattern covaried with the number of elements.

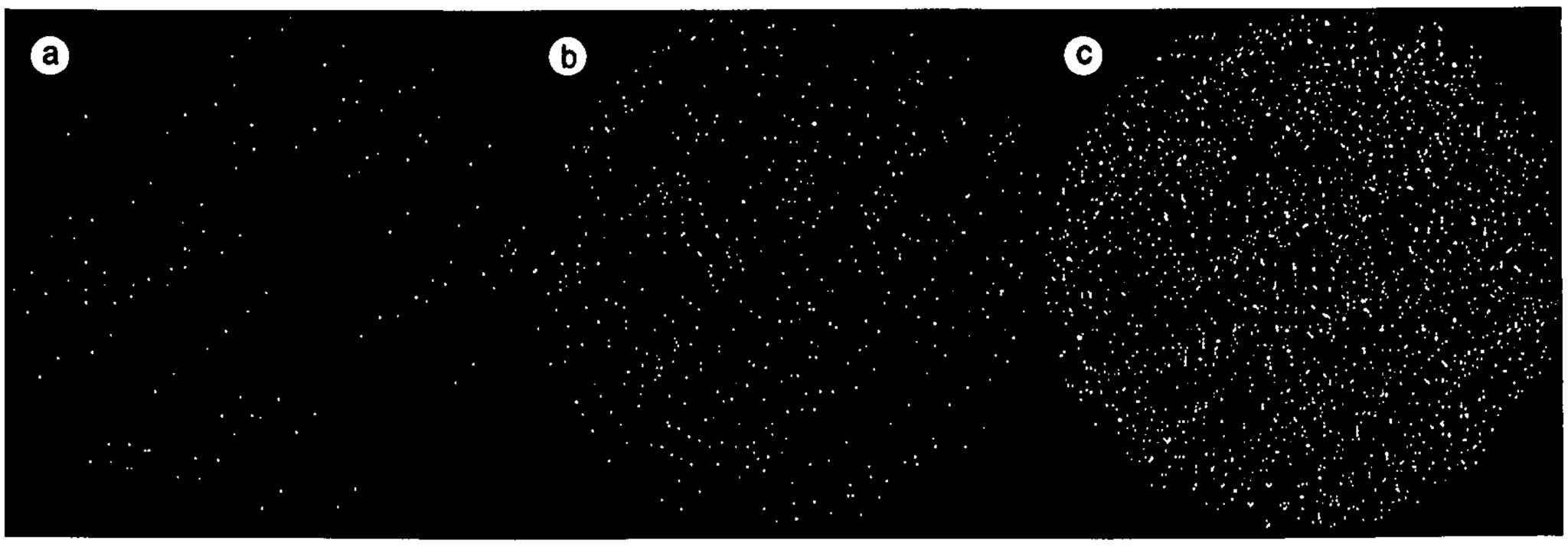


Figure 4.10. Examples of the stimuli used in Experiment 4.3. The patterns shown contain (a) 64, (b) 256, and (c) 1024 dipoles.

4.7.2 Stimuli

All stimuli were Glass patterns with a dipole separation of 8 arc min., a value sufficiently large that any effects of neighbours should become apparent as density is increased. No dipole orientation variance was added. The number of dipoles in each pattern was varied from 8 to 1024 in one octave steps. Examples are shown in Figure 4.10. Since constant field size and viewing distance were employed, dot density varied from 1.68 to 215 elements/degree^2.

4.7.3 Results

Threshold offsets for the mean orientation judgement as a function of the number of dipoles are shown for three subjects in Figure 4.11. The accuracy of judging orientation is slightly poorer for very sparse patterns, but rapidly improves with increasing number of elements and performance asymptotes for patterns containing 32-64 elements. There also appears to be a slight dip in the function for all three subjects around 32-64 dipoles. The basic pattern of the data shows that there is little effect of stimulus density over about 64 dipoles. These data are in accord with performance on a structure detection task, reported in Jenkins (1983).

That subjects are relatively insensitive to pattern density seems to be inconsistent with models based on neighbourhood matching. All three subjects' performance is as good with patterns containing 1024 elements as with those containing 64. These patterns have, respectively, an average of 0.5 and 8.0 dots lying closer to each dot than its correspondent. This is in agreement with data from Maloney *et al.* (1987) which showed that structure versus no-structure judgements were possible when dots had more than six other dots

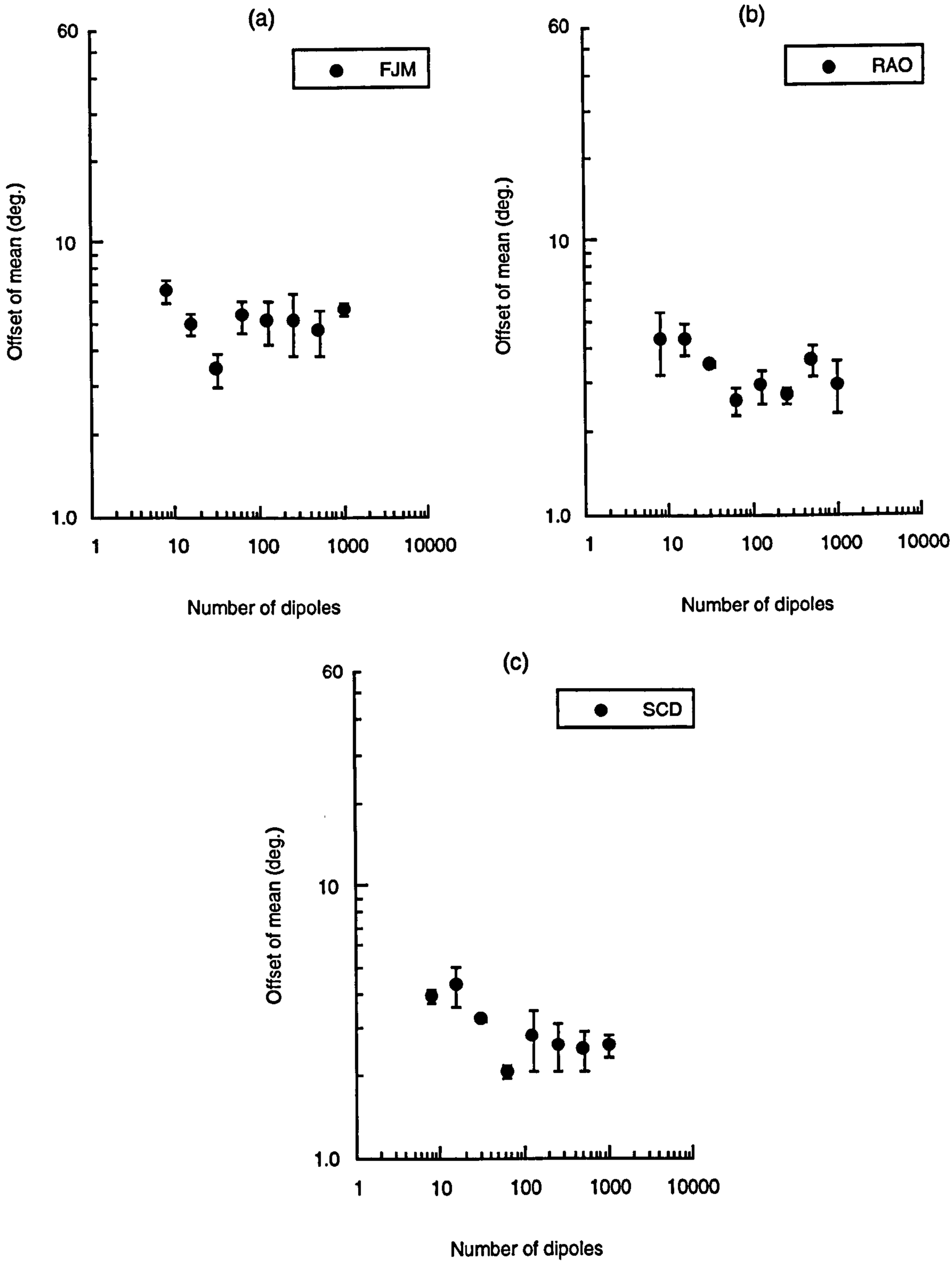


Figure 4.11. Threshold offset for judgement of the mean orientation of Glass patterns as a function of the number of dipoles in the texture.

closer to them than their correspondent.

4.8 Conclusions

To summarise the preceding three experiments, the stimulus parameters were as follows:

	mean orientation	*number*	*length*	*orientation s.d.*
Exp. 4.1	90°	512	1.41-32.0 arc min.	8°
Exp. 4.2	90°	512	8.0 arc min.	1.41°-32.0°
Exp. 4.3	90°	8-1024	8.0 arc min.	8°

The main results are as follows:

- Estimating the mean orientation of a Glass pattern becomes easier as dipole separation is increased up to a critical separation, of around 5 arc min., beyond which performance rapidly deteriorates. For line textures there is a consistent improvement with increasing line length.

- Local orientation variance has little effect on judging the mean orientation of line textures or Glass patterns until a standard deviation of around 8° is reached. A similar pattern of deterioration is observed with line and dipole textures, except that performance with Glass patterns is uniformly poorer.

- Subjects are highly accurate at performing mean orientation judgements with thresholds which asymptote at around 1.5° for line textures, and 2.5° for Glass patterns.

- The accuracy of judging the mean orientation of a Glass pattern, within the limits tested, appears to be largely independent of the number of dipoles used.

The following chapter examines the implications of these findings for models of local orientation extraction, and compares the predictions of four such models for the tasks described.

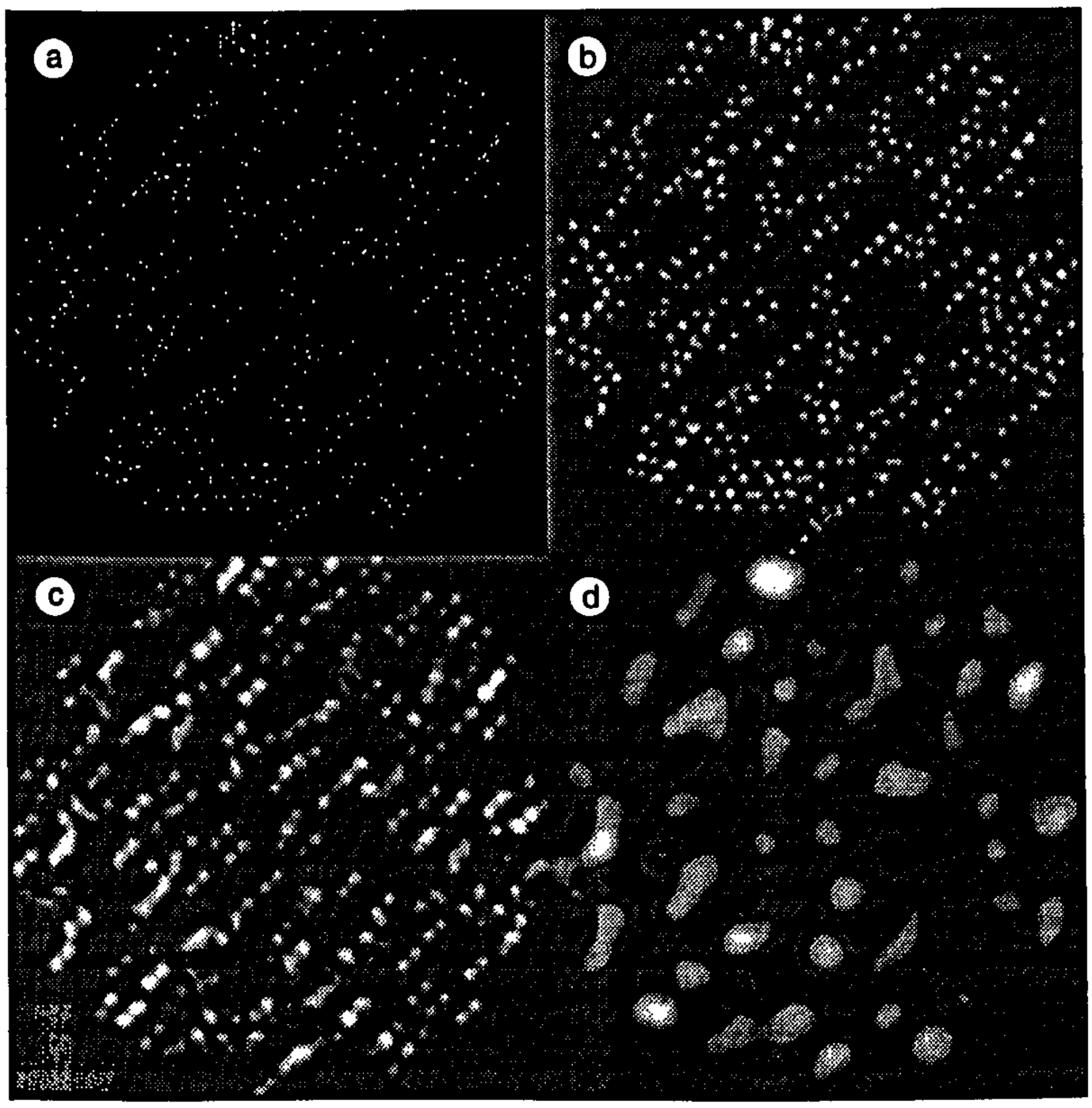
5 | Modelling of mean orientation judgements

5.1 Overview

In this chapter four models for extracting local orientation are described: symbolic matching of tokens (Stevens, 1978), isotropic (Laplacian-of-Gaussian) filtering, oriented (Difference-of-Gaussian) filtering and "adaptive" oriented filtering. The performance of these models was compared to human performance on the three tasks described in the previous chapter. Since mean orientation judgements place clear *statistical* constraints on the observer, it was possible to devise a simple method for determining the free parameter of the spatial filtering models, the correct spatial scale of analysis, based on minimisation of local texel orientation variation. Versions of both the isotropic and oriented filtering models incorporating this criterion were tested.

The simulations demonstrate that no one filter can explain data from the three tasks considered. Results suggest that only a model based on oriented filtering gives sufficiently accurate estimates of local orientation to account for subjects' performance with particularly noisy patterns. It is concluded that an adaptive filtering model, operating at a scale which minimises texel orientation variance, is an effective system for extracting local estimates of image flow. Neither the Laplacian-of-Gaussian model nor the symbolic matching scheme perform sufficiently accurately to account for human performance on the tasks considered.

Sections 5.2-5.4 of this chapter describe the texel isolation models and Section 5.5 the details of the simulations performed. Section 5.6 final section examines the simulation results.




```
BLOBF    Area      Mass          Centroid          Sd.          Error   Axis
DATA
BLOBD(   2498.00   -291.07   -12.84   125.58   29.95   5.26   0.08    3.438 )
BLOBD(   2034.00   207.32    -16.53   101.52   23.32   5.63   0.13    2.934 )
BLOBD(   2280.00   177.74    0.36     63.28    42.81   4.56   0.06    0.842 )
:        :         :         :        :        :       :      :       :
```

Figure 5.1. (a) Glass pattern with mean orientation of 45° and dipole separation of 8 pixels. (b-d) Laplacian-of-Gaussian filtered versions of (a), thresholded at ±1 standard deviation from the mean grey level. Filters have s.d.s of (b) 1 pixel, (c) 4 pixels (d) 16 pixels. The zero-bounded blobs are characterised using a sentence-based description, an example of which is shown in the bottom part of the figure.

5.2 Isotropic filtering model

A model for texel extraction using Laplacian-of-Gaussian ($\nabla^2 G$) filtering was implemented. It had five stages:

- Filtering with $\nabla^2 G$ at a single spatial scale.
- Thresholding to remove low responses.
- Description of resultant blobs using an image description scheme (Watt, 1991b).

- Selection of arithmetic mean of all blob orientations (weighted by blob mass and aspect ratio).

- Psychophysical decision.

The $\nabla^2 G$ is defined as:

$$\nabla^2 G(x, y, \sigma) = \frac{1}{\sigma^2}\left(1 - \frac{x^2 + y^2}{\sigma^2}\right) e^{-(x^2+y^2)/2\sigma^2}$$

The model was run using a range of filter scales: σ = 1.00, 1.41, 2.00, 2.82, 4.00, 5.66, 8.00, 11.31, 16.00 and 22.3 pixels. For a viewing distance of 2.0 metres, σ ranges from 0.60 arc min. to 6.68 arc min. in half octave steps. The filtering process is illustrated in Figure 5.1.

Following convolution the filtered outputs were half-wave rectified. The resultant blobs (e.g. Figure 5.1) have been proposed as the basic perceptual primitives of the raw primal sketch, rather than a zero-crossing based representation (Watt, 1988). This view is supported from psychophysical tasks such as edge blur discrimination (Watt and Morgan, 1983b), edge location (Watt and Morgan, 1985), and vernier acuity (Watt and Morgan, 1984).

The use of blobs from the thresholded output of the $\nabla^2 G$ as texture primitives has been proposed elsewhere (Voorhees and Poggio, 1987). The major difference between the model proposed in Voorhees and Poggio (1987) and the one implemented here is firstly, that here no gain control of the image preceded filtering, and, secondly, that thresholding took place at $\pm$1.0 standard deviation from the mean grey level (rather than Voorhees and Poggio's (1987) selection of threshold using histogramed local gradients). Generally the setting of the threshold (within reasonable limits) is not critical and pilot simulations, using thresholds from 0.5 to 4.0 standard deviations, revealed that a threshold of anywhere between 0.75 and 2.0 standard deviations is optimal for estimating local orientation. For this reason one would expect the results of the Voorhees and Poggio (1987) model to be similar to those of the model implemented here.

The zero-bounded regions in the image were then characterised using the image description algorithm of Watt (1991b). This produces a compact "sentence" describing each blob in terms of principal axis, centroid position, area, etc. Two features were actually used in calculating the mean orientation of the set: blob orientation and orientation reliability (details of the calculation of these attributes are given in Chapter 2, Section 2.3.3).

The mean orientation at each scale, $\bar{\theta}_\sigma$, was calculated using Equation 2.4 from Chapter 2. However, because translational patterns were being used, the average was estimated using *all* blobs in an image. Because unimodal orientation distributions were used in all of the experiments the basic models did not incorporate any form of orientation filtering. Finally, the anti-clockwise versus clockwise decision was made by classifying mean orientations between 0° and 90° as clockwise of vertical, and between 90° and 180° as anti-clockwise of vertical.

The model described so far only uses a single spatial scale of filter for performing the task. Given that the human visual system has a range of filter sizes available, it would be desirable to have a model which *automatically* selects the correct filter size from trial to trial. To this end an automatic version of the $\nabla^2 G$ model was implemented which incorporated a scale selection criterion. This operates by calculating the orientation reliability of the sentence-based description at each scale. Assuming that a model of mean orientation estimation will attempt to maximise confidence in its estimate then it should select the spatial scale at which orientation variability is minimal. Field orientation variability (Ψ_σ) was calculated using Equation 2.4.3.

In the automatic model the mean orientation, $\bar{\theta}_\sigma$, which had the lowest associated Ψ_σ was used to make the psychophysical decision. In the case of the pattern shown in Figure 5.1, the image shown in part (c) has minimal texel orientation variation and would be selected as the correct scale to make an estimate of local orientation.

5.3 Oriented filtering models

Motivated by the presence of cells in V1 which are not only sensitive to the spatial scale of a pattern, but to also to its orientation, a number of models have been proposed for deriving local orientation estimates using oriented filtering. Zucker (1982) has proposed a model which estimates image flow using the identity of the most locally active oriented Difference-of-Gaussian filters, in conjunction with a relaxation algorithm which maximises orientation consistency (i.e. collinearity) within a neighbourhood.

The two models described in this section also use DoG filtering but differ fundamentally from the Zucker (1982) model. Firstly, they do not use the *identity* of the most locally active filter to estimate orientation; instead the filter output is used to derive a symbolic description, from which orientation estimates are made. Secondly, both models do not incorporate an iterative post-filtering stage. Instead the intention was to test the

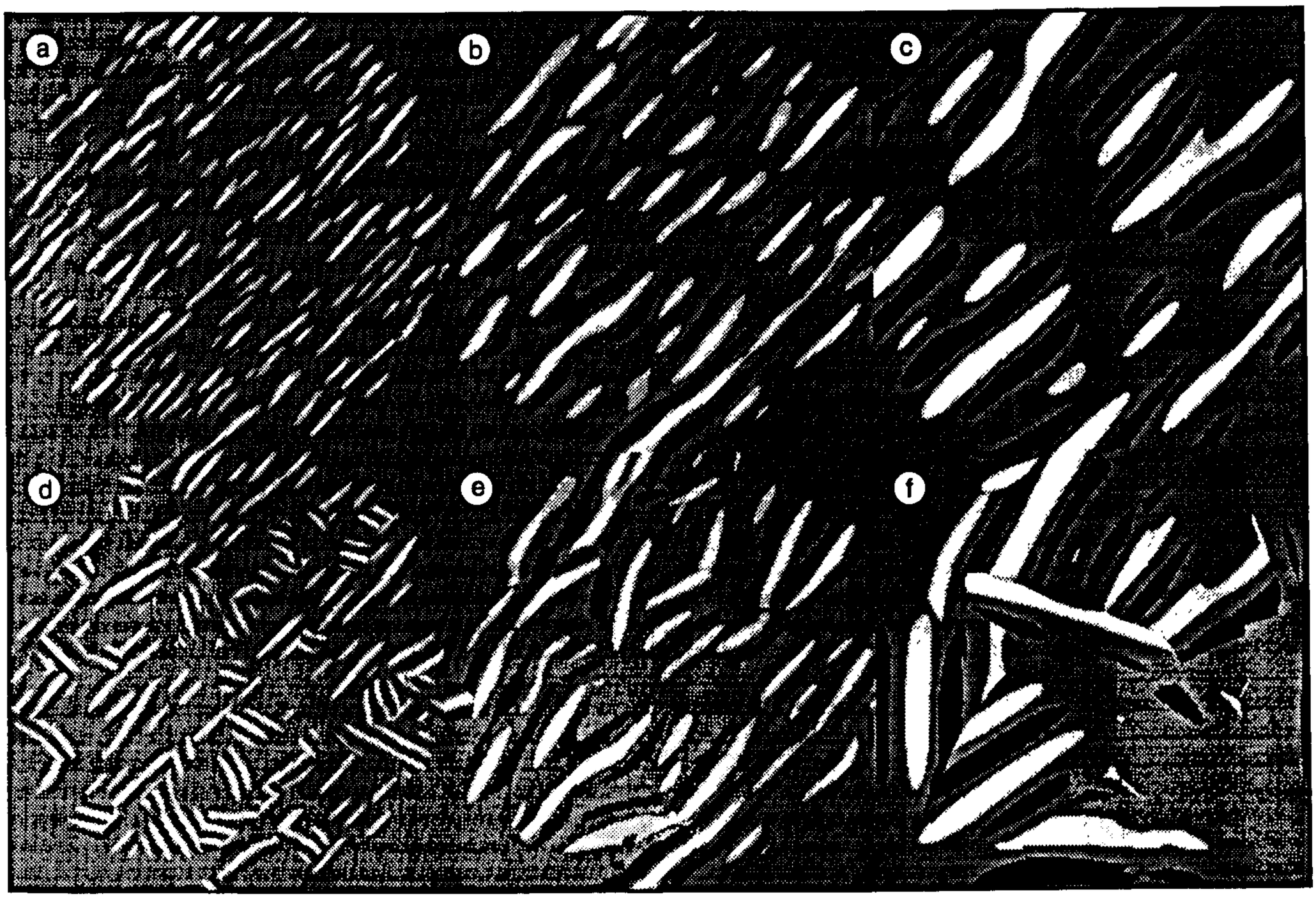


Figure 5.2. (a-c) Glass pattern filtered with 45° DoGs with $\sigma = 2, 4$ and 8 pixels, and thresholded at ± 1 standard deviation from the mean grey level (d-f) As above but using local selection of the most active filter, across orientation, at each scale. The zero-bounded blobs in both these models are characterised using a sentence-based description, to derive an estimate of mean orientation.

accuracy of orientation estimates measured using either single DoG filters or a scheme that ensures local orientation consistency by the addition of an adaptive component to the convolution stage.

The first oriented filtering models operates on the output of a single Difference-of-Gaussian filter, centred on the orientation which patterns varied around (i.e. the vertical). This model therefore assumes prior knowledge of the pattern orientation. These filters are composed of a DoG in the x-direction multiplied by a Gaussian in the y-direction:

$$W(x_t, y_t) = \left(e^{-x_t^2/2\sigma^2} - (1/2.23)e^{-x_t^2/2(2.23\sigma)^2}\right) e^{-y_t^2/2(3\sigma)^2} \tag{5.1}$$

where σ refers to the standard deviation of the positive Gaussian function. x_t and y_t are coordinates rotated by angle ϕ:

$$x_t = x\cos\phi + y\sin\phi$$

$$y_t = y\cos\phi - x\sin\phi$$

The ratio of the amplitudes of the positive and negative parts of the DoG and the aspect ratio of the filter are based on those derived by Wilson and co-workers using a variety of psychophysical paradigms (Wilson and Gelb, 1984; Phillips and Wilson, 1983). A range of filter sizes were employed with σ varying from 1.0 to $8\sqrt{2}$ (1.23 - 19.6 arc min.) in half-octave steps (all centred on 90°).

Output from filters was thresholded, characterised using the image description algorithm of Watt (1991b) and a mass-weighted estimate of mean orientation made (in the same way as in the $\nabla^2 G$ model). The operation of oriented filters, at different spatial scales, on a Glass pattern is illustrated in Figure 5.2a-c. Note that the blobs derived are all highly elongated in the direction of the filter orientation.

The problem with using the output of DoG filters at a single scale is how to deal with more complex images that contain orientation changes across the image. What is required is a scheme for integrating filter outputs across orientation. The second oriented filtering model uses *adaptive filtering* to perform the integration of orientated filter outputs at a single scale (the model is explained in more depth in Chapter 2). The model incorporates a pixel-by-pixel selection of the output of the most active DoG filter (across all orientations at one spatial scale). Twelve orientations of DoG were used at each spatial scale ranging from 0° to 165° in 15° steps. The same range of DoG scales was used as for the single DoG

model. Applying this model to a typical Glass pattern again produces highly elongated blobs (Figure 5.2d-f).

Exactly the same scale selection criterion was used as in the $\nabla^2 G$ model to select a sentence-based description of the adaptively filtered output at one spatial scale, i.e. mean minimum orientation variance of the blob descriptions. In the example images shown in Figure 5.2e shows the scale of adaptive filtering selected using this method. This criterion could not be applied to the single DoG model, because there was insufficient variation of blob orientation variance across scale.

5.4 Stevens' algorithm

A model described in Stevens (1978), based on token matching, was implemented. The model calculates all possible pairings of a single token to other tokens within a surrounding neighbourhood, and all possible pairings of those tokens in a similar sized regions surrounding them. All matches have a corresponding orientation which is histogramed (weighted by the proximity of the matched to the original feature). The most frequently occurring match orientation is used to select the correct correspondent for the original feature.

Proximity weighting in Stevens' model is relative to the neighbourhood size. The weighting of a virtual line's contribution to a local orientation histogram is either 1, 2/3 or 1/3 depending on whether neighbourings dots are less than 1/4, less than 1/2 or greater than 1/2 a neighbourhood radius apart, respectively.

Given a set of weighted orientations, a local histogram is constructed in order to compute the direction of local parallelism and thereby determine each dot's correspondent. Smoothing of the local histogram is performed using relatively small numbers of "buckets", and the peak orientation selected. Finally, the corresponding dot is determined by selecting the virtual line whose orientation most closely matches the peak orientation. If no line can be found within 15° of the peak, no solution is returned for the dot.

The model as described has three parameters which will determine its performance: the neighbourhood size, the proximity weighting function, and the degree of smoothing of local orientation histograms. Stevens claims that human performance is matched by using the proximity weighting scheme described and local orientation histograms with 18 buckets. He claims that the limiting factor on the models performance is the neighbourhood size. Figure 5.3 shows the effect of varying neighbourhood size on the performance

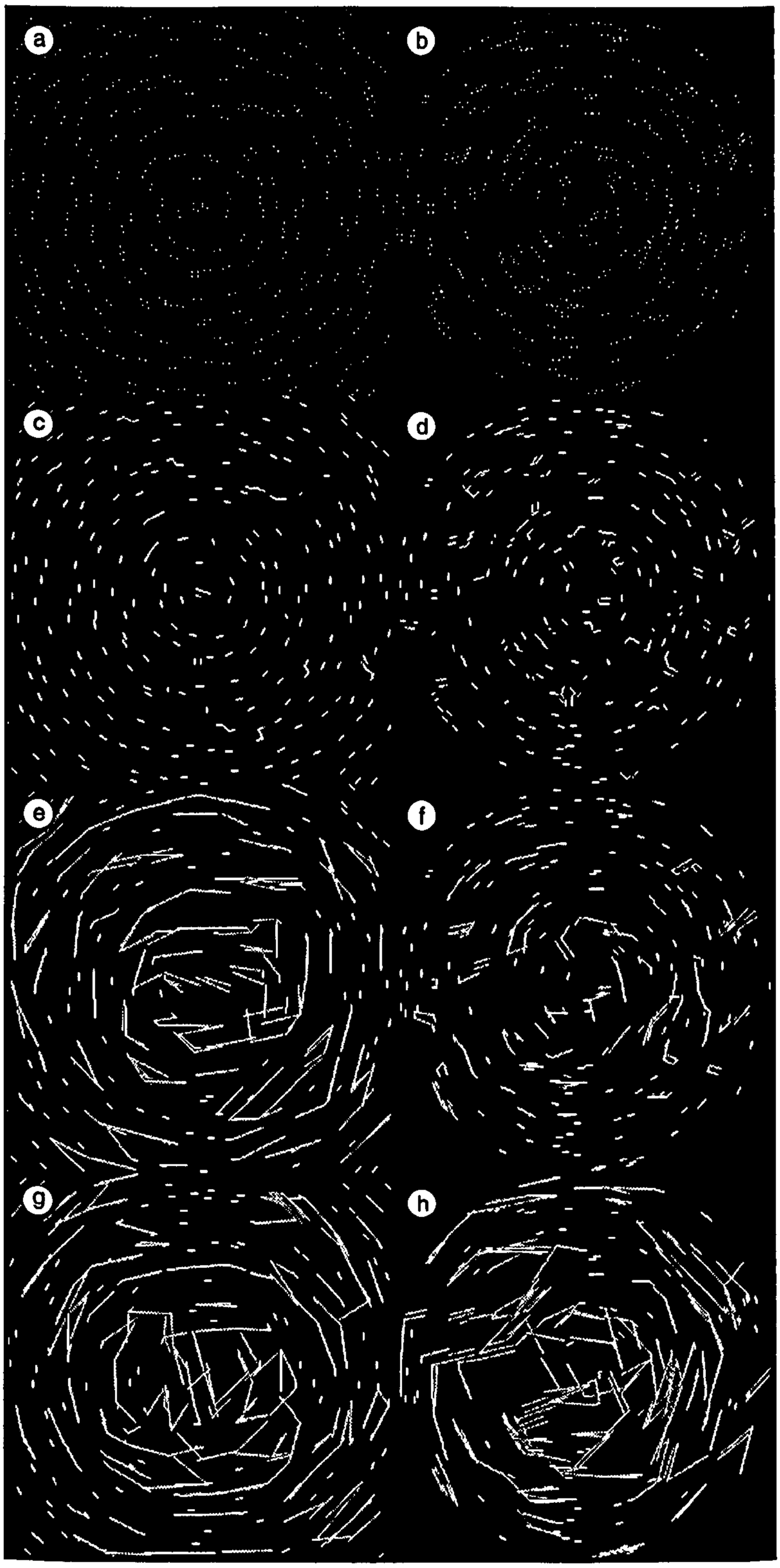


Figure 5.3. The effect of neighbourhood size on Stevens' (1978) algorithm. (a, b) Rotational Glass patterns (274 dots, dipole length = 5 pixels) with original dot set distributed on a perturbed grid (a) and uniformly randomly positioned (b). The three flow fields beneath each Glass pattern have been derived using Stevens' algorithm with different sized neighbourhoods. Neighbourhoods have radii of (c, d) 7 pixels, (e, f) 15 pixels and (g, h) 30 pixels.

of the algorithm on two rotational patterns. The first corresponds to the types of patterns considered in Stevens (1978): dipoles are spatially distributed over a perturbed grid. The second pattern is made up of randomly spatially distributed dipoles. Note that fields extracted from the second pattern are uniformly worse.

In the simulations described, two methods of setting the neighbourhood size were examined. The first set the radius equal to the dot separation, i.e. the optimal size for discounting unmatched dots. Stevens (1978) claims that, since subjects cannot see structure when more than two or three dots lie closer to a dot than its correspondent, such a small region will not tend to give enough samples to allow the reliable extraction of a peak orientation. This conjecture was tested in the simulation which follows. The second method of setting neighbourhood size used the overall density of the pattern. Stevens claims that a neighbourhood which contains six or seven dots closely emulates human performance on the psychophysical tasks he describes.

5.5 Details of simulation

To compare the models' and subjects' performance, simulations of Experiments 4.1-3 were performed. The models described were used to estimate the mean orientation of each pattern and measures of performance, based on the accuracy of those estimates, were calculated.

Simulation of the tasks was undertaken using a procedure as close to the experimental conditions as possible. In order to generate a threshold mean orientation from the models, a method of constant stimuli was used. Sixty-four stimuli were generated at each of the same 17 stimulus levels as were adaptively sampled in the psychophysical experiments. Each stimulus image was processed using either of the three models described, and an estimate of the mean orientation extracted. The value of this mean was used to make a decision if the stimulus pattern had a mean orientation which was clockwise or anti-clockwise of vertical, and the outcome of this decision was recorded for each model. Given a set of probabilities of the model making a correct judgement, as a function of the size of the cue, the standard deviation of this psychometric function was calculated using the same system as was used for data from the human subjects. This procedure was performed as a function of the independent variable of each experiment.

For models incorporating a scale selection criterion, the scale of analysis was chosen using the first 6 stimuli at the beginning of a run, and that filter size employed throughout

that run. Thus, each time that the independent variable (stimulus length, for example) was changed, this choice was made by running the full models over 6 example stimuli and using the mean spatial scale that the application of the criterion produced. The neighbourhood size parameter of the Stevens model was set by hand at the beginning of each run.

5.6 Simulation Results

5.6.1 Oriented filter models

The result of the simulation of the mean orientation judgement as a function of dipole length, using the two oriented filter models, is shown in Figure 5.4. Filled symbols represent the mean performance of the three subjects from Experiment 4.1, fine lines the predictions of individual DoG filters, and the coarse line the prediction of the adaptive filtering model. It is clear that *no one DoG filter can explain subjects' performance on this task*. If filter size is too small, or large, compared to the separation of the dipoles, only uniformly poor estimates can be made of mean orientation. The performance of these filters declines suddenly as the length of dipoles exceeds the size of the excitatory zone of the oriented filter; there is a small range of lengths for which a particular filter is optimally tuned. Predictions from the adaptive filtering model, operating at a scale determined by minimising texel orientation variation, closely match the performance of subjects. Note that no fitting has been applied to the model predictions.

The mean performance of the three subjects from Experiment 4.2 is shown in Figure 5.5 along with predictions from the single DoG and adaptive filtering models. Results suggests that the output of single DoG filters show the same pattern of deterioration in estimates of mean orientation, as a function of local orientation variance, as the human subjects. Furthermore, the output of a single DoG with σ between 4.90 and 9.8 arc min. fits subjects' performance well. Predictions from the adaptive model, shown as the heavy line, again match subjects' performance closely. The adaptive filtering model behaves very much like a single DoG model in this condition.

Figure 5.6 shows the simulation results for the task from Experiment 4.3. Results are similar to those from the last simulation: a single DoG filter with a s.d. between 4.90 and 9.8 arc min. shows the same trend as human data, as does the adaptive filtering model.

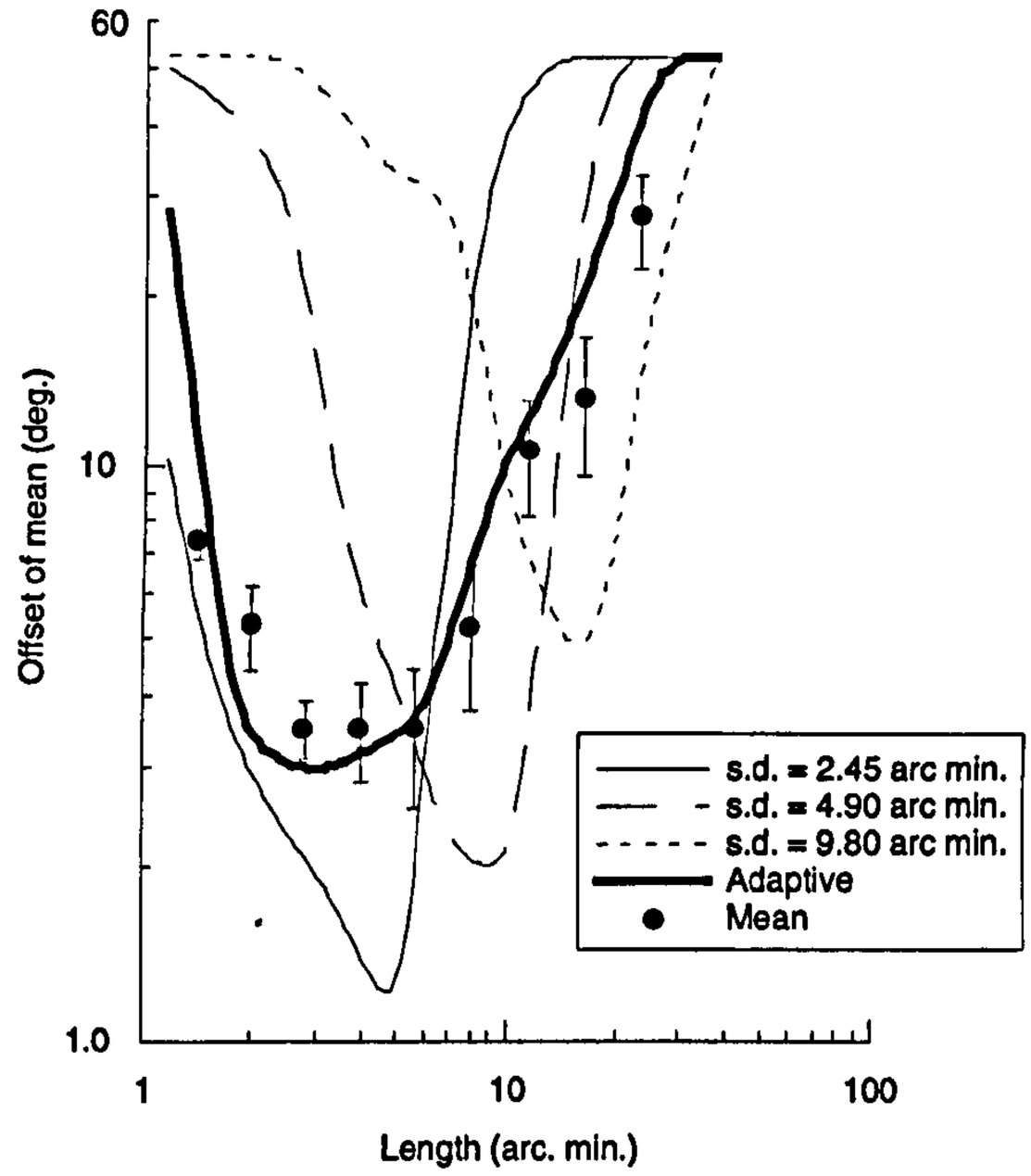


Figure 5.4. Threshold offset for DoG estimates of mean orientation of dipole textures, as a function of dipole length. Note that individual DoG filters are highly accurate at the task, over a small range of dipole lengths, but that the overall poorer performance of the adaptively filtered model closely matches human performance.

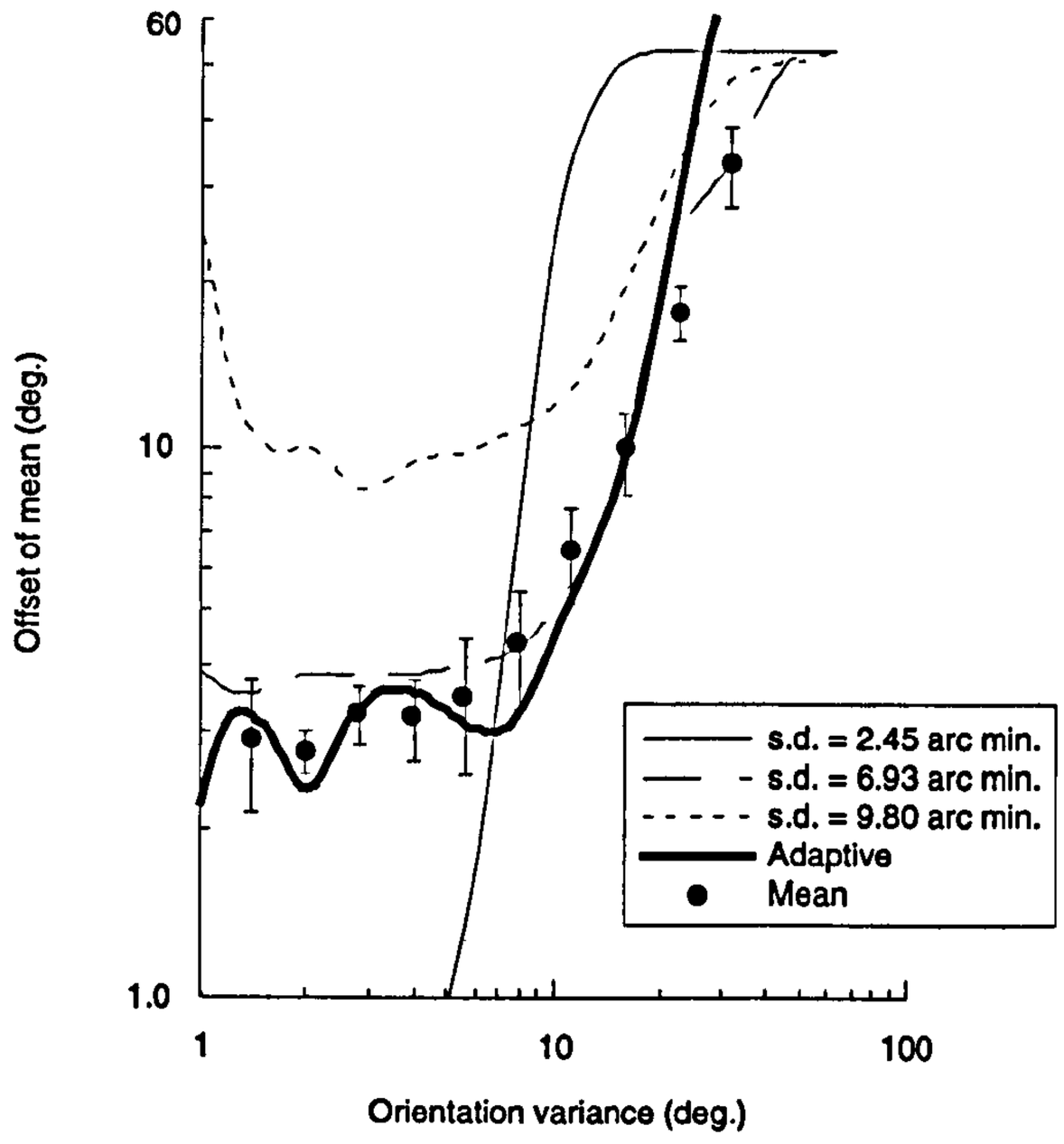


Figure 5.5. Threshold offset for DoG estimates of mean orientation of dipole textures, as a function of dipole orientation variance.

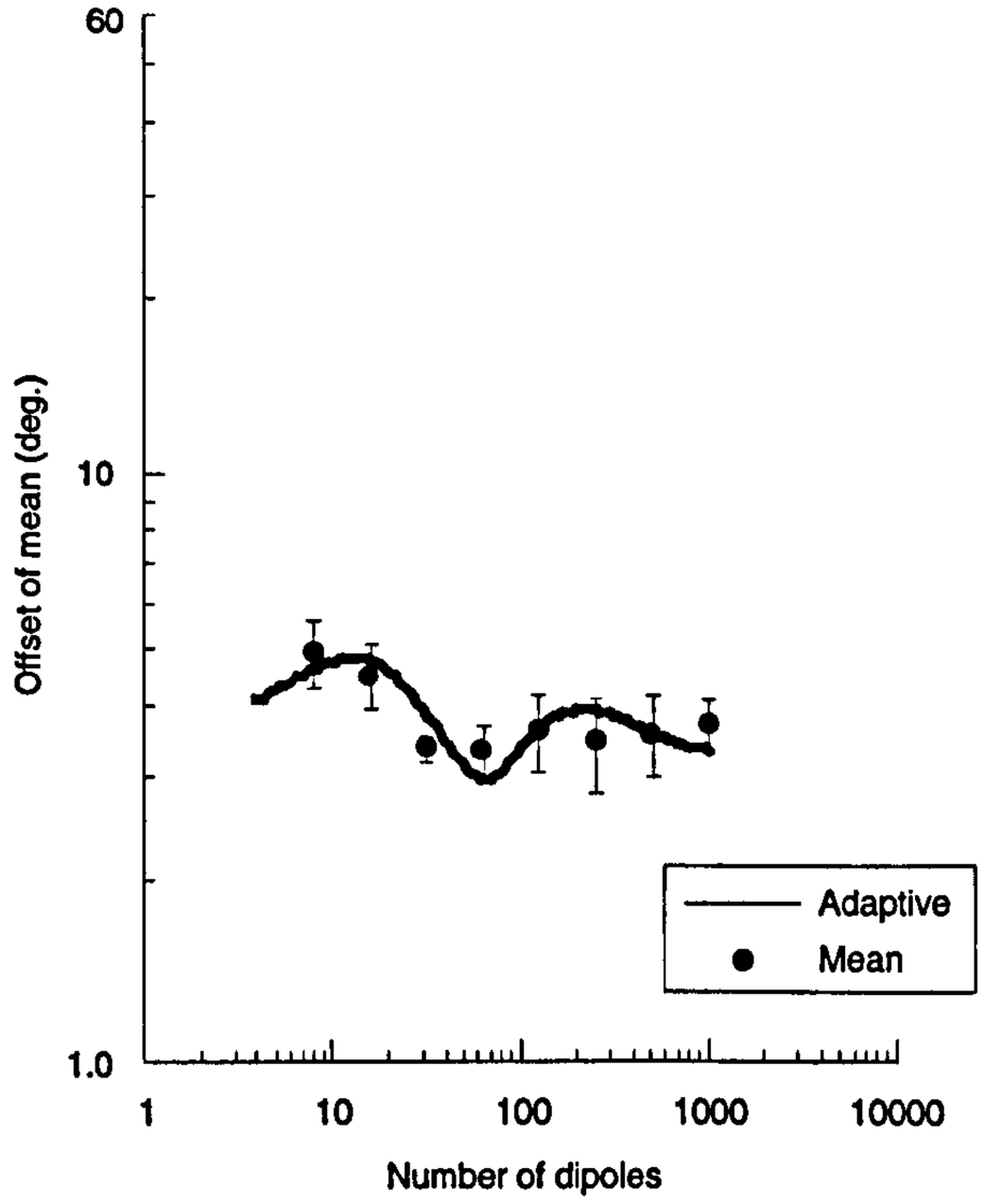


Figure 5.6. Threshold offset for DoG estimates of mean orientation of dipole textures, as a function of number of dipoles.

5.6.2 Isotropic filter model

Figure 5.7 shows the standard deviation of the mean orientation estimate as a function of dipole length for individual $\nabla^2 G$ filters. It is clear that for progressively larger dipole lengths, coarser scale filters give the best estimates of mean orientation. It is also apparent that no response from any one filter can explain the variation in subjects performance on this task. The solid line shows predictions of a model incorporating the scale selection criterion. Note that the overall pattern of responses is broadly similar to subjects. The primary difference is that the best performance of the model occurs around a narrow range of dipole separations about 2.5 arc min., whereas human performance is best around a broader range (2.5-5 arc min.) of separations.

Accuracy of the $\nabla^2 G$ model compared to human subjects, as a function of additional orientation jitter, is shown in Figure 5.8. Again the model produces the same pattern of responses as human observers, but this time does not approach their best performance on this task at low levels of orientation variance. Human subjects consistently achieve thresholds of around 3.0°, compared to the model whose best performance is around 5.0°. Although this is a small difference it is important because it suggests that a model based on the Laplacian-of-Gaussian *cannot explain the basic level of performance in this*

task. Deviations of a model from data which are due to the model exceeding human performance can be explained in terms of noise on the system. This type of deviation cannot. Figure 5.8 also shows that the critical level of orientation variance, beyond which performance deteriorates, for the model (around 20.0°) is greater than for human observers (around 8.0°).

It is possible that the failure of the model on this condition is due to the setting of certain parameters of the model, such as the degree of thresholding, the use of a single spatial scale, etc. To try and at least partially take this into account of this simulations of Experiment 4.2 were re-run, with three different levels of thresholding (0.5, 1.0 and 2.0 grey-level standard deviations), and incorporating four different levels of integration across scale (average across texels from ± 1 or ± 2 octaves of spatial scale). These variations did not produce any improvement in performance beyond that presented in Figure 5.8. Similarly, to ensure that background levels of orientation "noise" were not producing this poor performance, the simulation was re-run using Laplacian-of-Gaussian filtering of the orientation histograms, before extracting the centroid.Again, this produced no appreciable improvement on performance.

While this does not preclude the possibility some other way of treating the output of Laplacian-of-Gaussian might produce better results, it does at least suggest that the result is not an artifact of the setting of some individual variable within the model, as described.

Finally Figure 5.7 compares predictions from the $\nabla^2 G$ model and data from the density condition, Experiment 4.3. There is a reasonable match between human data and the predictions from the model incorporating automatic scale selection, although the basic level of performance of the model is again slightly higher than data.

In summary, this section has suggested that a model which uses $\nabla^2 G$ filtering, to extract and describe texture primitives, shows a similar pattern of results to the data from Experiments 4.1-4.3. Howeyer this model fails to explain the performance of subjects in Experiment 4.2.

5.6.3 Stevens' model

Predictions from the Stevens' (1978) model on the three tasks are compared to human performance in Figures 5.10, 5.11 and 5.12. Even though the predictions shown are for the model operating using the optimal neighbourhood size for the pattern, it performs

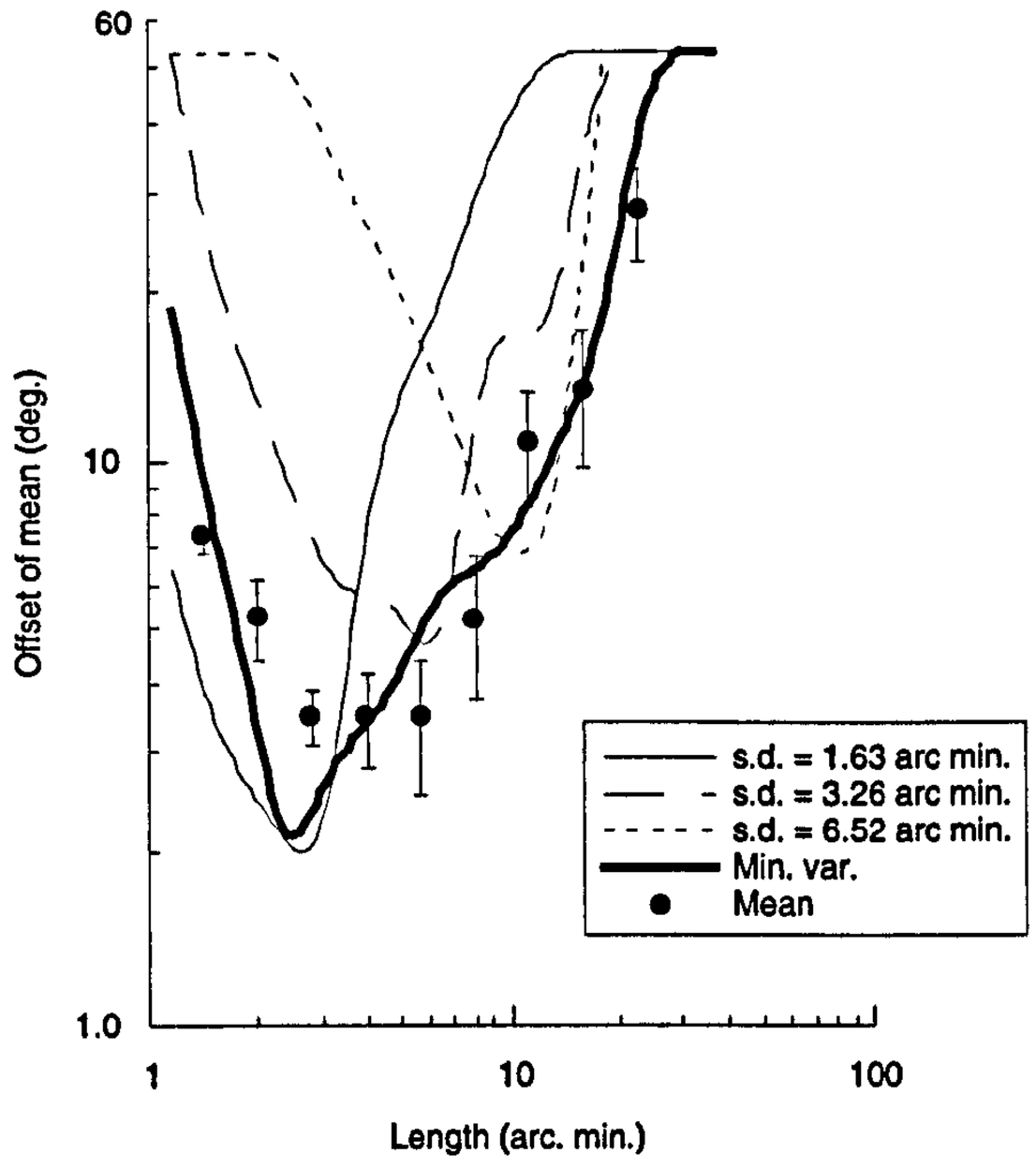


Figure 5.7. Threshold offset for $\nabla^2 G$ estimates of mean orientation of dipole textures, as a function of dipole length.

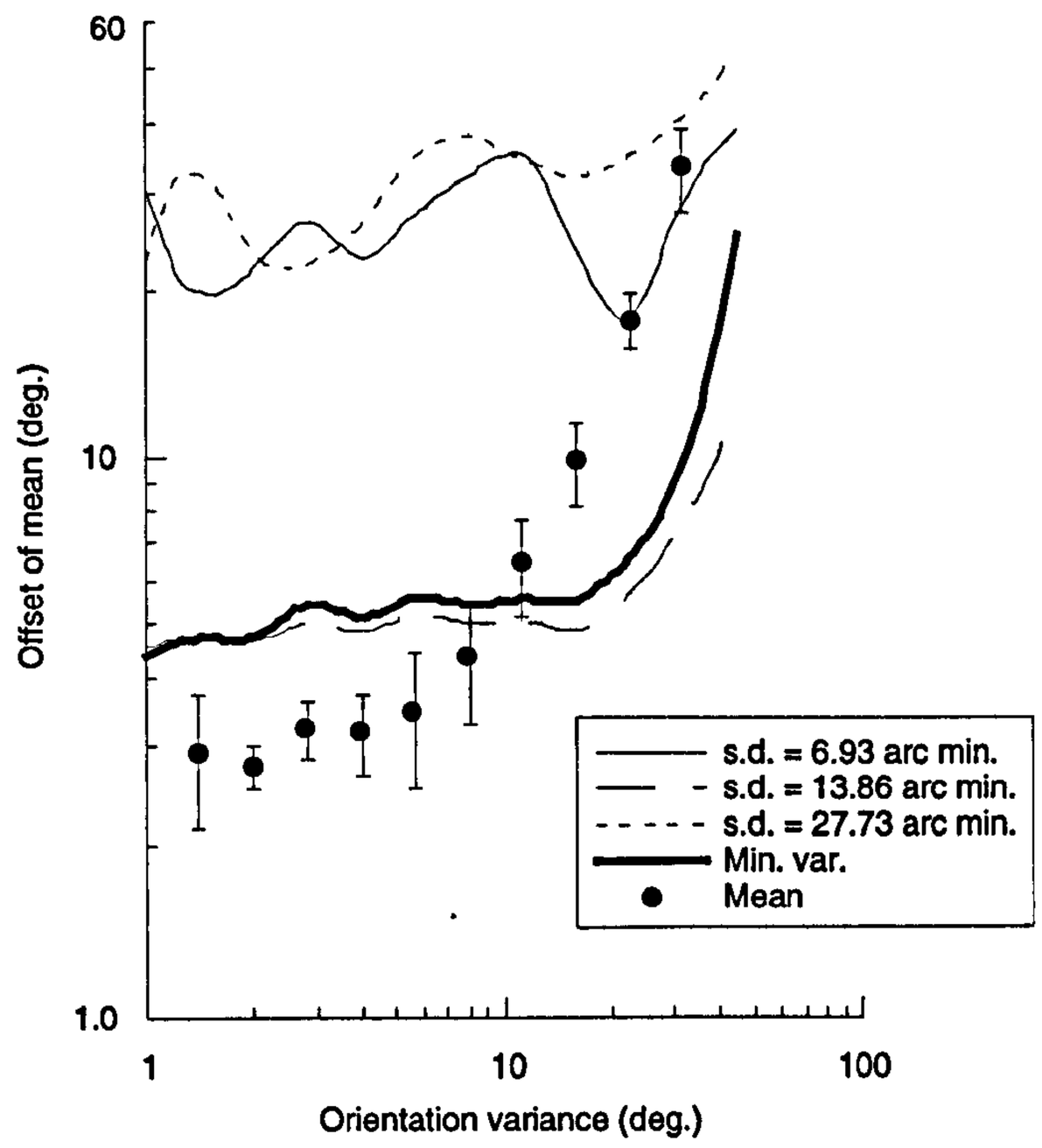


Figure 5.8. Threshold offset for $\nabla^2 G$ estimates of mean orientation of dipole textures, as a function of dipole orientation variance.

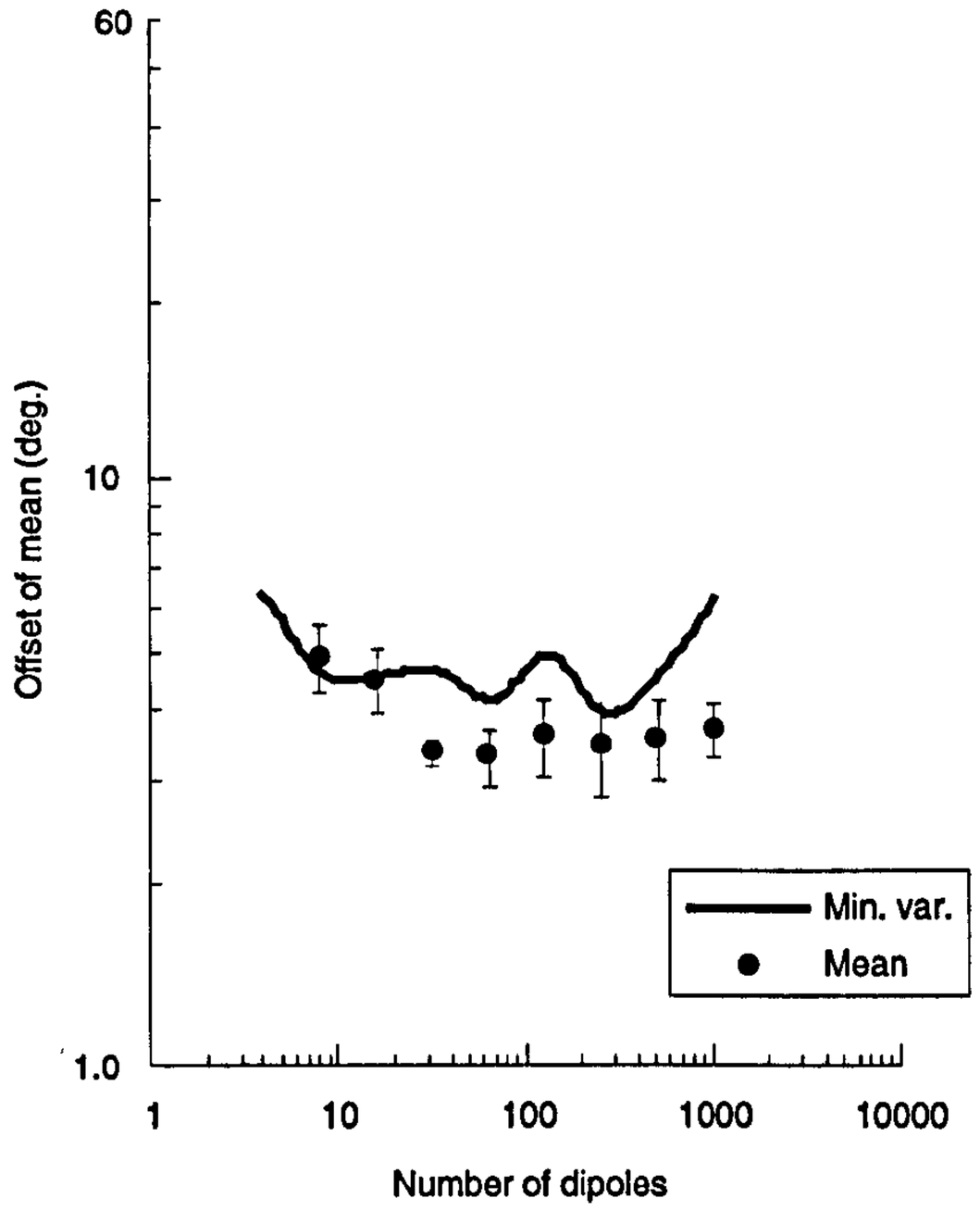


Figure 5.9. Threshold offset for $\nabla^2 G$ estimates of mean orientation of dipole textures, as a function of number of dipoles.

uniformly worse than human subjects and the other models. Predictions from the model using the pattern density to set neighbourhood size, are uniformly poor and are not presented. Given the way this model operates, this strongly suggests that in detecting structure in the Glass patterns used as stimuli in the last chapter, subjects make use of the *low spatial frequency information* which arises from accidental co-alignments of dipoles. Stevens' model can only use *individual* dipole orientations and when dipole separation is large in relation to dot density, as was the case in our experiments, matching of individual dipoles breaks down. It is concluded that this model cannot provide an explanation for the perception of structure in Glass patterns in any but the most limited of cases.

5.7 Conclusions

Weighted chi-squares of the fit of the three models to the subject data, from the three tasks using dipole textures, are given below. Asterisks indicate that the goodness-of-fit was significant to at least the 0.05 level.

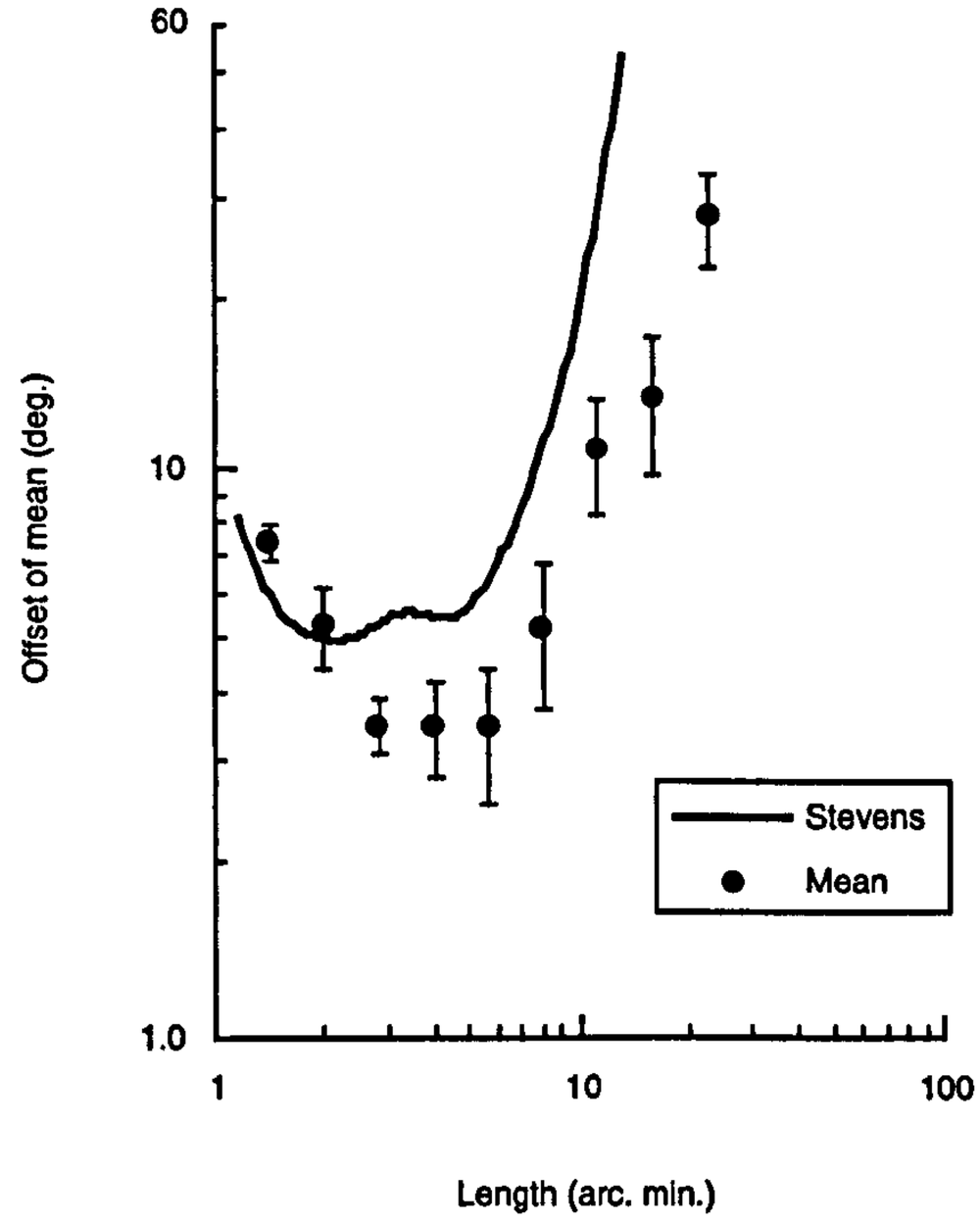


Figure 5.10. Threshold offset for estimates of mean orientation of dipole textures, from Stevens' algorithm, as a function of dipole length.

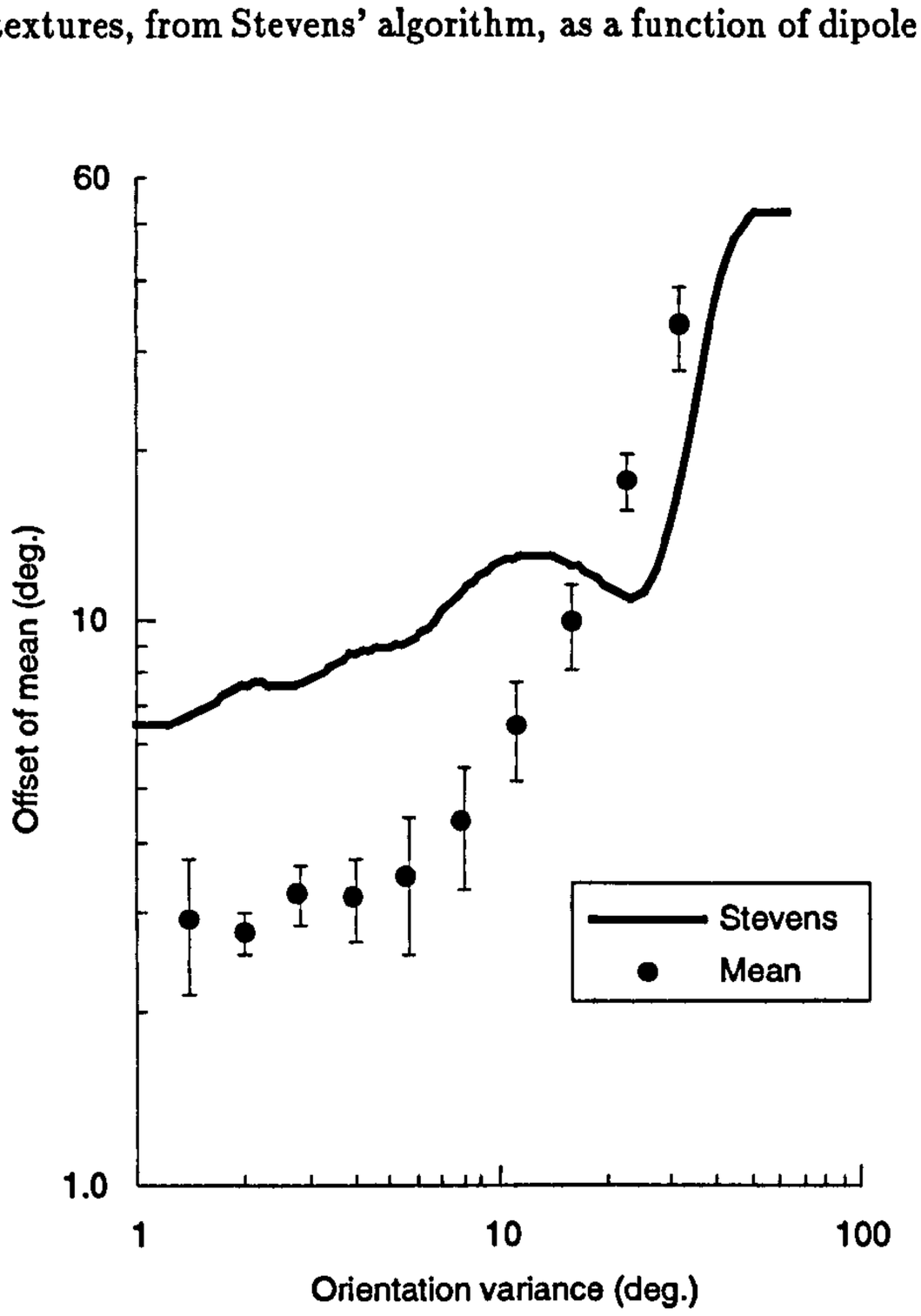


Figure 5.11. Threshold offset for estimates of mean orientation of dipole textures, from Stevens' algorithm, as a function of dipole orientation variance.

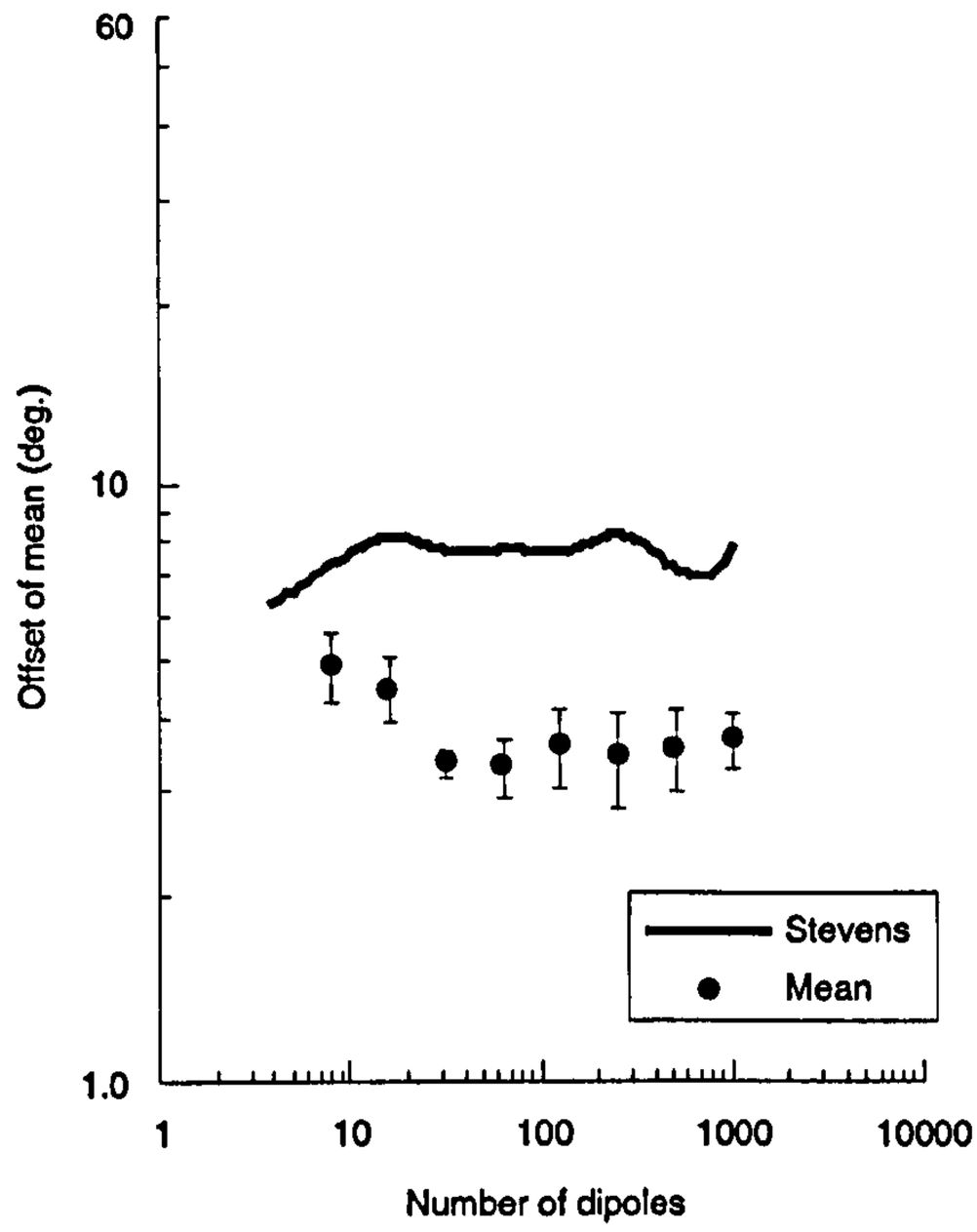


Figure 5.12. Threshold offset for estimates of mean orientation of dipole textures, from Stevens' algorithm, as a function of the number of dipoles.

	Adaptive DoG	*Auto.* $\nabla^2 G$	*Stevens*
Exp. 4.1	0.90*	1.65*	7.71
Exp. 4.2	1.31*	4.87	5.31
Exp. 4.3	0.15*	2.02*	4.83

Stevens' model fails to achieve a significant fit to the data in any of the conditions. The $\nabla^2 G$ model produces significant fits to the data from the orientation variability and dipole density conditions but the goodness-of-fit is not significant in the length condition. This failure, along with failures of the Stevens' model, are important because the model performance in all of these cases was *worse* than that of human subjects. The adaptive filtering model produces significant fits in all conditions and, on the grounds of parsimony, is accepted as the best model of subjects performance in the experiments.

Grouping stimuli, such as Glass patterns, allow the study of feature extraction in texture flows because they isolate orientation as a useful source of information at a narrow range of spatial scales. Subjects' judgement of the mean orientation of these patterns as a function of orientation jitter, dipole length and dipole density suggests that the visual system accurately selects the correct filter size which gives the best estimate of mean orientation. It was suggested that, by means of the nature of a mean orientation judgement,

an appropriate spatial scale of analysis would minimise the variance of resultant features. When combined with an estimate of local orientation, measured using oriented filters, this criterion proved to account adequately for the data from the three conditions.

6 | Globally organised textures

6.1 Overview

The orientation structure of natural images is often much more complex than the simple translational orientation fields considered so far. This chapter examines how structure is derived from textures that contain systematic orientation change in space (*globally* organised textures).

In order to understand the representation of orientation in globally organised textures fields, most previous work has employed "structure versus no-structure" judgements, using stimuli such as Glass patterns. In this chapter the effects of two types of noise on such a judgement were investigated: the addition of uncorrelated dots, and the presence of local orientation variance. A measure of mean local orientation variance, Ψ, operating on the output of the adaptive filtering model, was found to signal structure effectively and was used (a) to select the spatial scale of analysis and (b) to discriminate structured texture from noise. A simulation of the task described in Maloney *et al.* (1987), using Ψ, was compared both to human performance and to an alternative model which measured the consistency of local orientation measures with prior knowledge of the underlying transformation (a *template matching* model). Both models produce reasonable fits to human data although (a) the local variance model does not perform well with low density textures and (b) the template model consistently exceeds human performance. Psychophysical data are also presented for tasks using textures with added orientation variance. Here the maximum orientation variance which allows discrimination of structure from noise was measured. Simulations, using the local variance and template matching models, again demonstrates the sufficiency of the models as explanations for human performance.

As well as preserving orientation structure due to an underlying transformation, the presence of *local discontinuities* in the orientation field (which can signal occlusion, surface

boundaries, etc.) should be signalled. A second task investigated this and was a partial replication of one presented by Hel Or and Zucker (1989): the detection of a patch of unstructured dots within a translational flow field. The unstructured field produces a change in mean local orientation variance which is again sufficient to explain human performance on this task. This calls into question the rather complex explanation, based on curvature, given for the task by Hel Or and Zucker (1989). It also suggests that other texture segmentation tasks might be achieved by subjects making estimates of the local smoothness of the field rather than detection of the borders of the disparate region.

6.2 Introduction

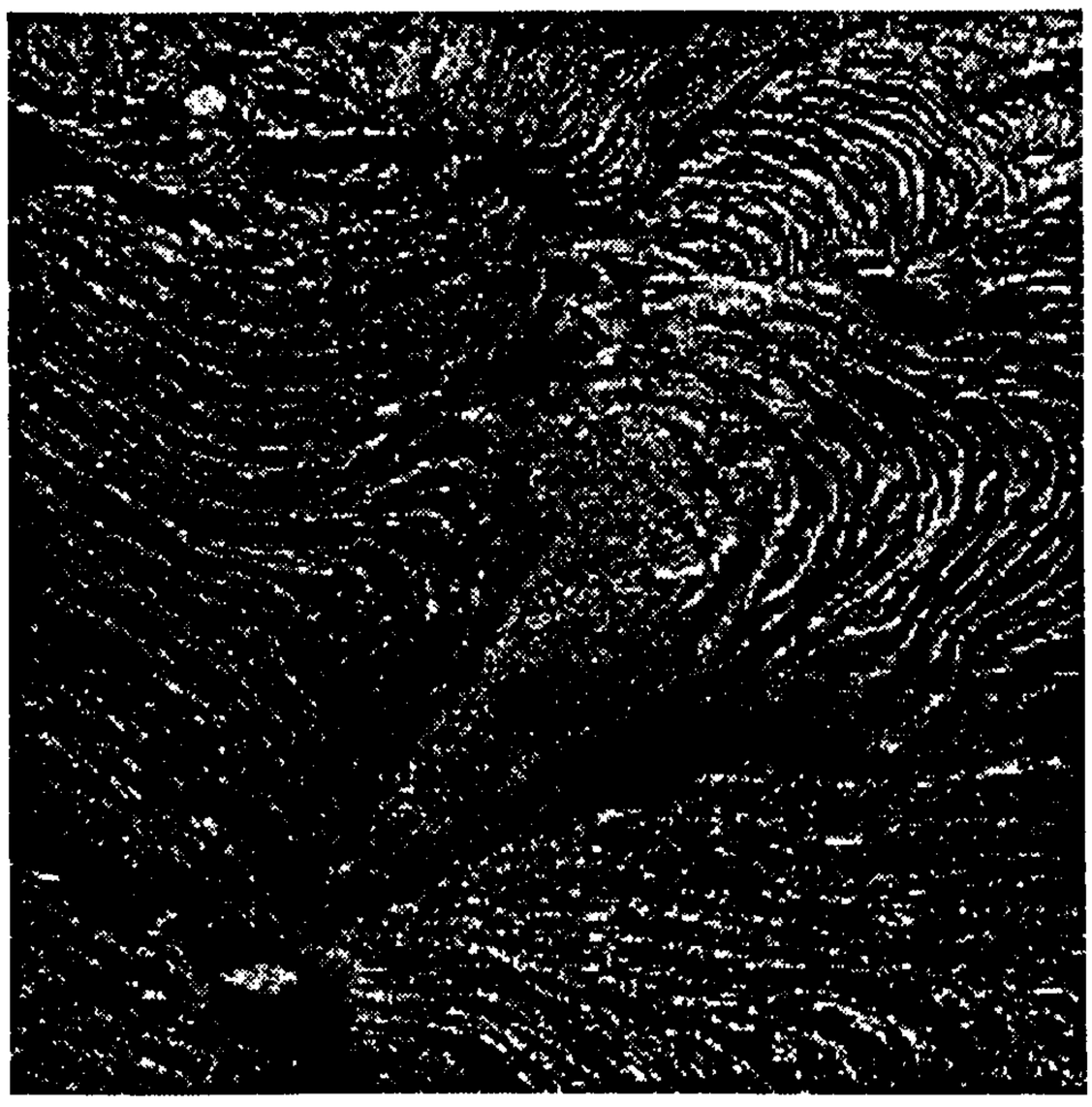


Figure 6.1. Image of lava flow illustrating complex orientation structure.

The model described in Chapter 2 was proposed as a system for extracting and representing orientation information from natural images. This model has been used to generate predictions of human performance primarily on the psychophysical judgement of mean orientation, because of the constraints this task places on what information subjects use from texture. However the orientation structure of natural images is often a great deal more complex than the translational orientation fields used in these experiments. This is because the physical processes underlying the formation of natural texture are subject to complex interactions.

Consider the image of a lava flow shown in Figure 6.1. The patterns of ripples in the rock are due to differences in the rate of flow of the waves of magma at the time of formation. These differences may be due to a large number of factors: irregularities in the underlying surface, variation in the forces propelling the magma, etc. As the magma cools, these factors and their interactions cause the complex flow shown in the texture. Because of the liquid nature of the medium these interactions do not result in discrete and independent texture elements but instead in extended contours (following the ridges in the surface). The viscosity of the magma and the rate of cooling determines the maximum rate of change of contour orientation over the texture and in this case both

are sufficiently high to result in an orientation field that does not appear to randomly change across the field. Instead orientation information within local regions of the field appears to be relatively uni-directional. This suggests a constraint on what is meant by a globally organised orientation field: that the texels within local regions of the texture will all have similar orientations. More formally we would expect that the *orientation variance of elements in local regions will be low for a smooth orientation field.* It is this constraint of local field smoothness that was the motivation behind the minimisation of Ψ, the mean local orientation variance (calculated from a subset of texels within local patches of the texture), as a general scale selection criterion in Chapter 2. The minimum texel orientation variance, used previously as a criterion for filter scale selection (Chapter 5), is a special case of this. When dealing with a translational field the orientation variance estimate may be calculated using texels from the entire pattern.

The model uses the local *mean* orientation within a region to calculate flow direction; the local orientation variance constraint may simply be viewed as a maximisation of the system's confidence in these estimates. Averaging of texel orientation estimates in space to derive their mean is necessary for a number of reasons. Firstly in dense textures local orientation information may be corrupted by noise - local averaging will counteract this. Secondly in sparser texture fields we still perceive uniform flow which suggests that some form of averaging of measures is occurring (Hel Or and Zucker, 1989). The amount of smoothing to be applied, i.e. the number of regions used, will depend on what is to be done with the orientation field derived. It is proposed that two forms of structure in orientation fields should be preserved and that this adequately constrains the region size used for averaging local orientation measures.

Figure 6.1 illustrates these two forms of structure. The first is the change of local orientation in space. Calculation of orientation must occur over a number of local regions, or the visual system would smooth over the local structure of this pattern; local orientations derived by the model should reflect the change in contour orientation over the pattern. The second form of structure is *abrupt changes in orientation* in the texture, e.g. at the intersection of the smooth, wedge-shaped region in the centre of Figure 6.1, with the strongly oriented regions in the left and lower portions. This sudden change in texture attributes, and orientation in particular, that occurs at the boundary of two textures has been very intensively studied, both in the field of psychophysics (for review see Bergen, 1991) and computational theory (for review see Ajuha and Schachter, 1982).

There are two principal reasons for this. The first is due to the influence of the early psychophysical work of Julesz. The second is due to Marr, who proposed that a map of local surface discontinuities, estimated from a variety of visual cues such as colour, texture, luminance, etc., was the basis of the 2.5D sketch. This is a heavily investigated aspect of texture perception. In general abrupt changes in texture statistics, in analogy with abrupt changes in luminance, are thought to signal surface discontinuities. How true this is has not yet been demonstrated for natural images.

The preservation of discontinuities versus the calculation of representative local flow direction constrains the scale of averaging of orientation measures within the texture field. Too fine scale and the resultant field will not reflect correct local orientation because of noise. Too coarse and local orientation discontinuities will be smoothed out. To summarise it is proposed that the spatial averaging of local orientation estimates is constrained to maximise confidence in its estimates of the local mean orientation whilst retaining discontinuities in the orientation field.

Section 6.3 investigates the preservation of global structure in texture. The estimation of degree of structure using the mean local orientation variance measure (Ψ) is described. Sections 6.4-6.5 consider the use of this measure, both as a scale selection criterion and a discriminator of structure-from-noise, in the presence of two forms of visual noise added to globally organised flow textures. Sections 6.4 considers an experiment performed by Maloney *et al.* (1987), examining the effect of adding uncorrelated dots to globally organised Glass patterns. A simulation of their structure detection task is reported, using the adaptive filtering model with scale selection using minimisation of mean local orientation variation. Two variations on this model are compared. The first discriminates between structure and noise patterns by selecting the pattern with the lowest mean local orientation variation and the second by maximising the goodness-of-fit of the orientation fields from both textures to an idealised or template flow field.

Experiment 6.1 presents psychophysical data from a "structure" versus "no-structure" task, examining the effects of local orientation variance. The amount of orientation variance at which this task is performed at threshold was determined for Glass patterns containing various dipole lengths. This task is also simulated using the variance and template matched models.

In Section 6.6 of this chapter, the preservation of orientation discontinuities is investigated, through consideration of the problem of detecting anomalous patches embedded

in image flows. An extension of a paradigm presented by Hel Or and Zucker (1989) is used, to confirm the dependence of minimum detectable patch size on the number of dots making up a Glass pattern element (*path length*; Hel Or and Zucker, 1989). The data from this condition are simulated using the local variance and the template models, and an adapted version of the model operating on features calculated from local variance statistics.

6.3 Measuring the degree of structure

Because exclusively translational flow fields were studied in the last chapter, a simplified version of the adaptive filtering model was used. In order to determine the most appropriate spatial scale of analysis, an estimate of the smoothness of the orientation field was minimised; simply the standard deviation of all texel orientations in the pattern. However if we wish to estimate field smoothness for globally organised patterns (e.g. a rotational field) then a multi-local version of this estimate must be introduced. This estimate is defined in Equation 2.4.3 (Chapter 2) as:

$$\Psi = \frac{\sum_{(i,j)\in I_\sigma} \psi(i,j)\rho^2}{XY}$$

Where X and Y are the image dimensions, ρ_σ is the sampling rate, at scale σ, and $\psi(x,y)$ is a measure of the orientation variance in subregion (x,y):

$$\psi(x,y) = \frac{1}{\nu(x,y)-1} \sum_{(i,j)\in N(x,y)} \sin^2(\theta(i,j) - \bar{\theta}(x,y))$$

$\nu(x,y)$ is the number of blobs in each subregion. In the simulations that follow all parameters were set according to table 2.1.

However, given that what we are trying to measure using field smoothness is some measure of *structure*, an alternative might be to use any underlying knowledge we have of the orientation structure of the image. The simplest method is to calculate the deviation of local mean orientations $\bar{\theta}(x,y)$ from the known transformation at that position $(T_{x,y})$:

$$d(\sigma,x,y) = \frac{\sum\sum \sin^2(\bar{\theta}(x,y) - T(x,y))}{\rho_\sigma^2}$$

and to sum over all regions in the image:

$$D_\sigma = \sum_{(x,y)\in I_\sigma} d(\sigma,x,y)$$

$T(x, y)$ is the expected orientation within a subregion. For a rotation (centre, (x_c, y_c); radius, r_c), for example, the expected value $T_r((x, y)$ is defined as:

$$T_r(x, y) = \tan^{-1}\left(\frac{y_c - y}{x_c - x}\right) + 90°$$

for a dilation:

$$T_d(x, y) = \tan^{-1}\left(\frac{y_c - y}{x_c - x}\right)$$

and for a translation, angle A,

$$T_t(x, y) = A$$

The obvious disadvantage to this method is that the underlying transformation must be known in order to calculate $T(i, j)$. However, given that for the Maloney *et al.* (1987) study, considered in the next section, different types of patterns were not interleaved in a single run, it may be assumed that subjects had that information.

The remainder of this chapter presents psychophysical evidence relating to the effects of two forms of noise on our estimates of structure in flow patterns.

6.4 Experiment 6.1: Effect of uncorrelated dots on structure detection

This experiment was designed to apply the model described in Chapter 2 to globally organised stimuli and to compare its discrimination of structure in flow textures, to that of human observers.

Maloney *et al.* (1987) presented the first attempt to quantify the strength of the Glass pattern percept using signal detection theory. Unpaired randomly positioned dots were added to organised Glass patterns, and their discriminability from a pattern composed of randomly oriented dipoles was determined using a detection task. Data from Maloney *et al.* (1987) using a dilation of 18 arc min. and a rotation of 9 arc min. are the unfilled circles plotted in Figures 6.3 and 6.4, respectively. As might be expected, discriminability (d') decreases with increasing noise for all dot densities tested. Furthermore, there is a Weber law dependence of d' on the ratio of paired to unpaired dots. In this section we consider whether the model proposed for deriving orientation estimates from flow fields predicts such a dependence.

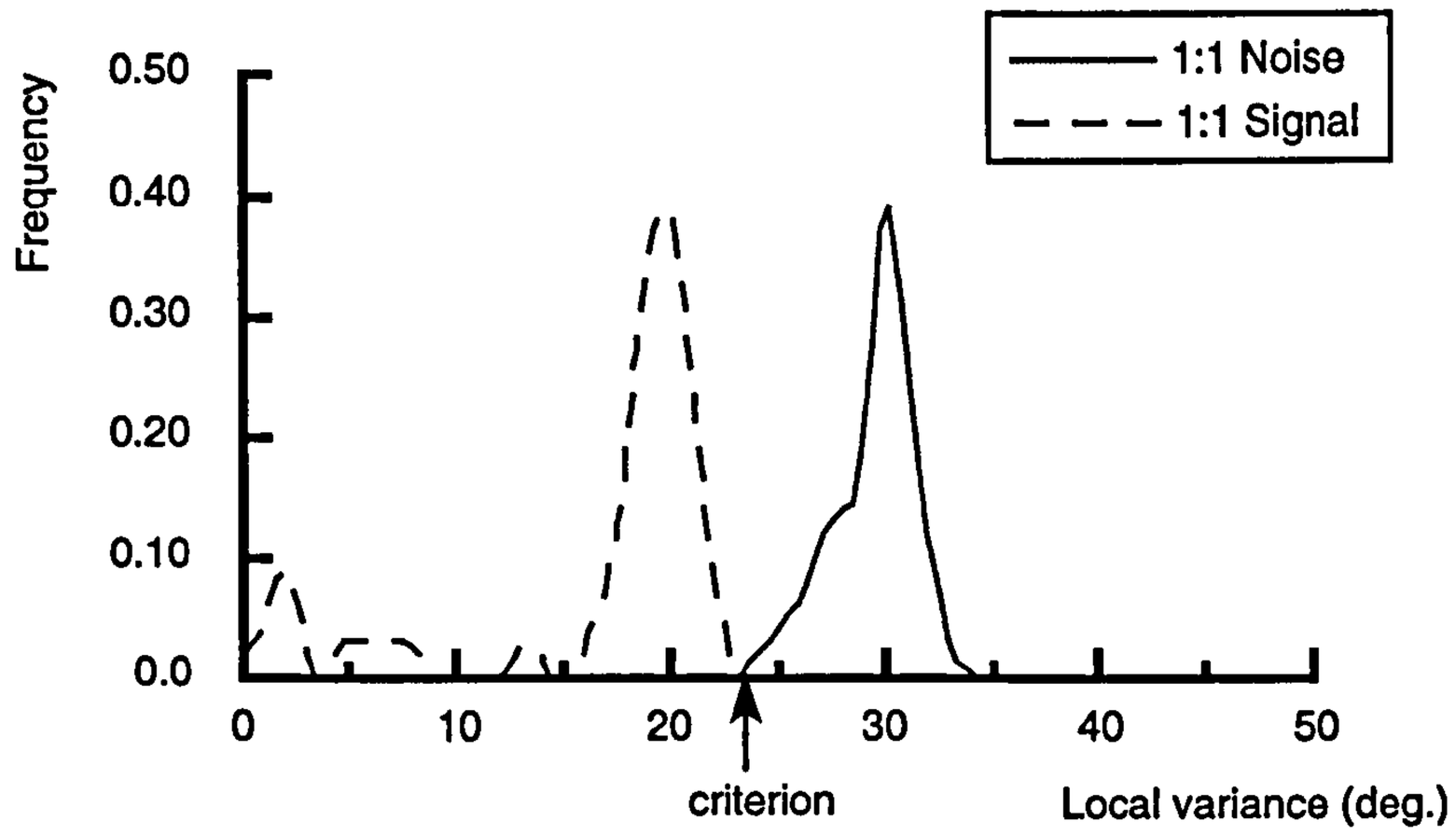


Figure 6.2. Distributions of the local variance cue for random and organised Glass patterns. The pattern used to generate this example was a 9 arc min. rotation containing 100 paired and 100 random dots. The distributions do not overlap and so, by placing the criterion at the mean of the two distribution means (as indicated), perfect signal detection performance is achieved.

6.4.1 Details of simulation

The adaptive filtering model, described in Chapter 2, was used to isolate texture elements from input patterns. The model operated at a single spatial scale, which was set at the beginning of the simulation according to the minimum local variance criterion. All other details of the model are identical to those described previously.

Given the local orientation based description of the flow field, two strategies were tested for deciding whether an input pattern was an organised or a random texture. The first, the *field smoothness* cue, is to use the same estimate of local mean variance (Ψ) that is used in selecting spatial scale. The local rate of change of orientation is expected to be slower for the organised than for the random patterns. The second cue tested was the *template matched* cue. This is the total difference between local estimates of orientation and the *expected* orientation, at that position, given a particular underlying transformation. Thus the second cue assumes that subjects have an accurate underlying representation of the idealised flow pattern.

Simulations were run using similar stimuli to those in two conditions from Maloney *et al.* (1987): rotations of 9 arc min. and translations of 18 arc min. For a pattern containing D dipoles and N noise dots, simulations were run at $D = 10, 25, 50, 100$ and 200 and $N = D, \sqrt{2}D, 2D, 2\sqrt{2}D ... 16\sqrt{2}\sqrt{2}D$. At each stimulus level, 32 organised and

random patterns were generated, and an estimate of the magnitude of the variance and template-matched cues calculated for the two patterns. These 32 values for each pattern were stored and used to calculate the discriminability, d', of the patterns, at each stimulus level (typical distributions of the variance cue are plotted in Figure 6.2).

The criterion (c) was set at the point that divides the means of the two sets (μ_1 and μ_2) in the ratio of the standard deviations of the sets (σ_1 and σ_2):

$$c = \mu_1 + \frac{\sigma_1}{\sigma_1 + \sigma_2}(\mu_2 - \mu_1)$$

One thousand random selections were made from the organised and random cue sets and each compared to this criterion. If the cue from an organised pattern exceeded the criterion the trial was classified as a hit. If the cue from a *random* pattern exceeded criterion, the trial was classified as a false alarm. Then d' was calculated from hr, the fraction of structured patterns labelled as structured and fa, the fraction of random patterns labelled as structured with:

$$d' = P^{-1}(\mathrm{hr}) + P^{-1}(\mathrm{fa})$$

where $P^{-1}(y)$ is the inverse function of the Gaussian probability function:

$$P(y) = \frac{1}{\sqrt{2\pi}} \int_0^y e^{(-x^2/2)} dx$$

Values of this function can be calculated to an arbitrary level of precision using an approximation to the incomplete Gamma function (given in Press *et al.*, 1992). However, if no false alarms occur, or if the hit rate is 100%, then the value of d' is meaningless and only related to the number of trials. To allow the assignment of some value to d' in these cases, the maximum hit rate permitted was 99% (i.e. hit rates greater than 99% were set to 99%) and the minimum false alarm rate permitted was 1%. In this way the maximum possible d' was approximately 4.653, which is the same as the maximum tabulated value in Swets (1964), the source used by Maloney *et al.* (1987) for calculating d'.

6.4.2 Simulation results and discussion

Results from the simulations are shown in Figure 6.3 (for rotational fields) and Figure 6.4 (for translational fields). The template model performs at least as well as human subjects in the ten conditions shown. As the total number of dipoles increases, however, this model displays a pronounced plateau around low signal-to-noise ratios (SNRs) indicating perfect

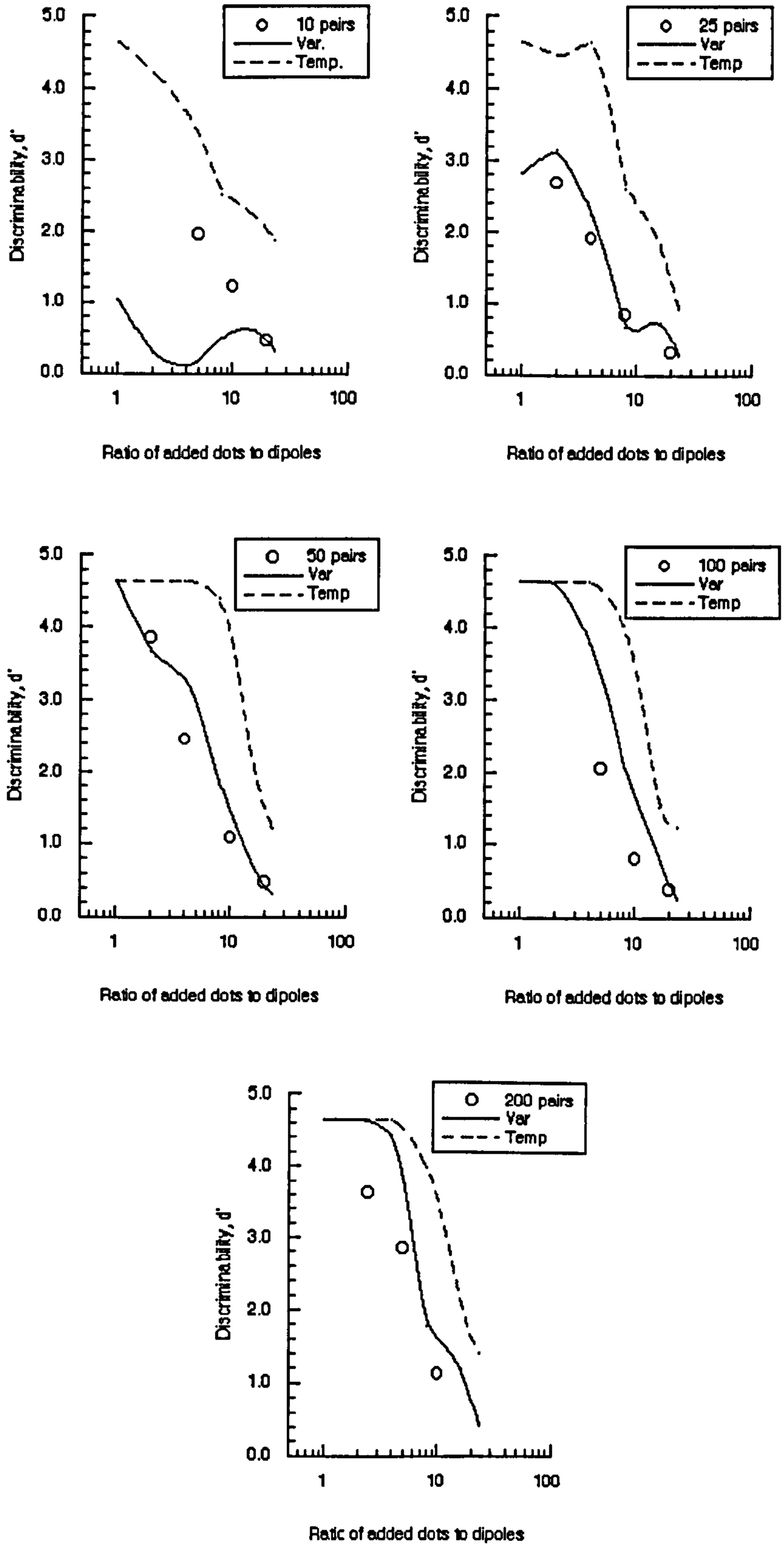


Figure 6.3. Discriminability of 9 arc min. rotational Glass patterns from random pattern by human observers (unfilled circles) and two models. The lines, marked "Temp" and "Var", show predictions from a field smoothness measure and a template match, respectively, both operating on the output of the adaptive filtering model.

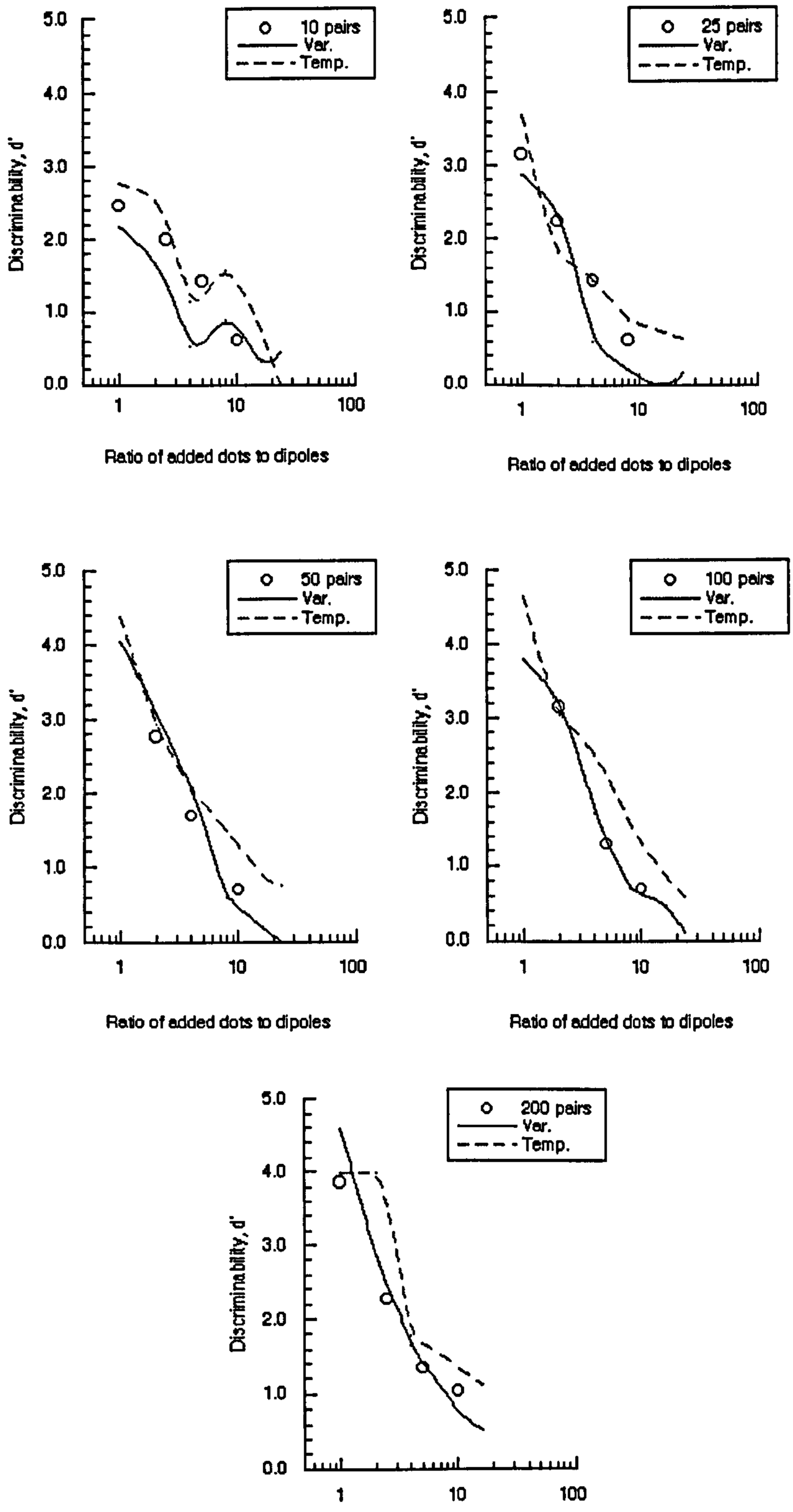


Figure 6.4. As Figure 6.3 but for 18 arc min. translational Glass patterns.

discrimination performance. The addition of some form of intrinsic noise on such a system would bring it into closer agreement with human data. Note that neither model shows a pure linear dependence on the ratio of added to paired dots - as predicted by Maloney *et al.* (1987). Both models plateau for high ratios of signal to noise. Although no such plateau is evident in data from Maloney *et al.* (1987), performance at the lowest ratio tested (1:1) was often around $d' = 4.0$ and is it assumed that at higher SNRs performance would have plateaued. The discrepancy between the template model and human data is greatest for the task using rotational fields, shown in Figure 6.3. This transformation used a dipole length of 9 arc min. compared to 18 arc min. for the translational field. This suggests that the template model is able to make the most of low uncertainty on *local* orientations to produce near perfect discrimination. The variance model on the other hand is, proportionally, less disrupted by uncertainty on local orientation. This makes sense given that the variance operator combines multiple orientation measures, giving some resistance to local orientation uncertainty but losing some information in the process.

Both the variance and template models show similar slopes to the data for virtually all densities tested. The variance model produces closer fits to human performance than the template model in most conditions (see Table 6.1), usually due to the extremely good performance of the template model. The agreement of both models and data is impressive given that there are no free variables used to fit predictions. However the variance model fails to achieve human levels of performance on very sparse rotations (Figure 6.3a) and translations (Figure 6.4a). There are a number of possible explanations for this failure. Firstly, the degree of smoothing used to calculate the flow field might not be sufficiently great, leading to "holes" in the orientation field and a deflation of the estimates of field smoothness. Secondly, the field might be too sparse for a local orientation variance measure to be of any use. In the limit, the template model can operate on one texel as, given knowledge of what transformation to expect, can the subject. The variance measure requires multiple orientation estimates, and if those measures are too sparsely distributed the measure will be useless.

In conclusion, whilst the superior accuracy of the template matching measure is clear, the local variance model generally provides a good account of the results reported by Maloney *et al.* (1987) and correctly predicts an approximately Weber law relationship between additive noise and number of dipoles. The template model performs impressively

		10 els	25 els	50 els	100 els	200 els
Variance	Translation	0.701*	0.784	0.213*	0.0242*	0.220
	Rotation	1.77	0.319*	0.325*	0.979	0.371*
Template	Translation	0.532*	0.306*	0.451	0.824	0.298*
	Rotation	5.42	12.4	9.37	9.96	4.3

Table 6.1. Chi-squares for the fits of the variance and template models to the data from Maloney *et al.* (1987). ("els" - elements)

in all conditions producing perfect detection at SNRs up to 1:8.

It is possible that a system using knowledge of the underlying orientation is limited in the number of estimates that can be used, so that this strategy is computed less reliably for dense patterns than the variance cue. A clear prediction from this account is that a subject's *uncertainty* about the transformation underlying a texture should be most damaged at low texel densities.

6.5 Experiment 6.2: Effect of orientation variance on structure detection

The model described uses local orientation variance to determine the spatial scale and local region of integration. It would therefore seem sensible to compare its performance on texture flows, with noise added to local orientation, to that of human subjects. In the following experiment the effect of element length, and the addition of local orientation variance on the perception of structure in Glass patterns, were again examined. However, since the nature of the stimuli precluded the use of a mean orientation judgement, a structure versus no structure paradigm was used. Specifically the maximum amount of tolerable orientation variance was determined, using an adaptive method, as a function of element length. This method has been applied elsewhere using mesh-derived textures (Hallett, 1992), and vector patterns (Dodwell and Caelli, 1985).

6.5.1 Method

Subjects

Five subjects took part in the experiment. SCD, DW and SW had corrected-to-normal vision. SCD (the author) was experienced in psychophysical procedure, and was aware of the purpose of the experiment while all other subjects were naive to the purpose of the

experiment and relatively inexperienced with psychophysical procedures. A short training period (64 presentations) was undertaken before data collection began.

Procedure

Similar equipment to previous experiments was used for stimulus presentation and response collection. Viewing distance was again two metres and presentation time was 100 ms, to prevent saccading during stimulus presentation. All Glass patterns and noise patterns were composed of dipoles of length 8 arc min.

The subjects' task was a two interval, two alternative forced choice. In one, randomly selected interval the texture was a Glass pattern, in the other a texture composed of randomly oriented dipoles. The subject indicated which of the two textures was a Glass pattern by depressing one of two keys on the computer keyboard. The independent variable was the amount of orientation jitter placed on the orientations of the dipoles in the Glass pattern. APE adaptively sampled a range of jitters going from $\pm 90°$ (where discrimination from the noise texture will be at chance) to $\pm 0°$ in steps of 22.5°. Thus the threshold value indicates the jitter tolerable for 83% correct performance. Other details of the experimental procedure are identical to Experiments 4.1 and 4.2.

Stimuli

Details are similar to previous experiments, except that the orientation of dipoles was determined by one of three transformations: rotation, dilation or vertical translation (examples are shown in Figure 6.5). Note that only dipole *orientation* was determined by the transformation and not length, as would be expected if the dot pair were the result of the application of a pure rotation or dilation upon a random set of dots. If real transformations had been used, the presence of shorter elements near the focus of the pattern would have provided a cue to the presence of the Glass pattern. Whilst a similar length transformation to the random dipole pattern could have compensated for this, it is orientation which is the primary attribute of interest in this experiment and for this reason only the orientation components of the rotation and dilation transformations were used.

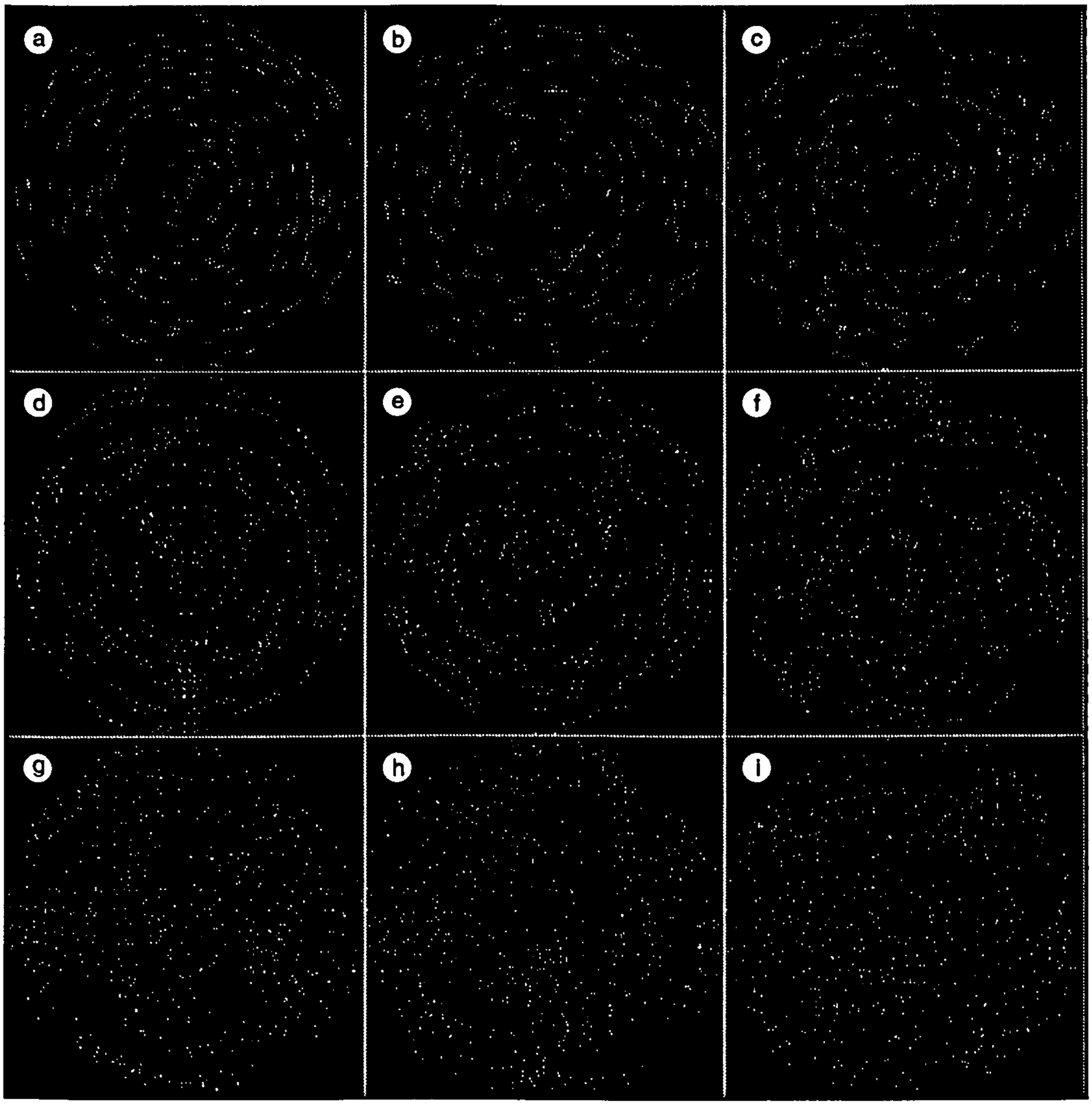


Figure 6.5. Examples of the stimuli used in Experiment 6.2 Each row shows three Glass patterns composed of 256 dipoles of the same length: (a-c) 4 pixels, (d-f) 8 pixels, (g-i) 16 pixels. Images in each row differ in the amount of orientation variation added to dipoles: (a, d, g) $\pm 0°$, (b, e, h) $\pm 22.5°$, (c, f, i) $\pm 45°$.

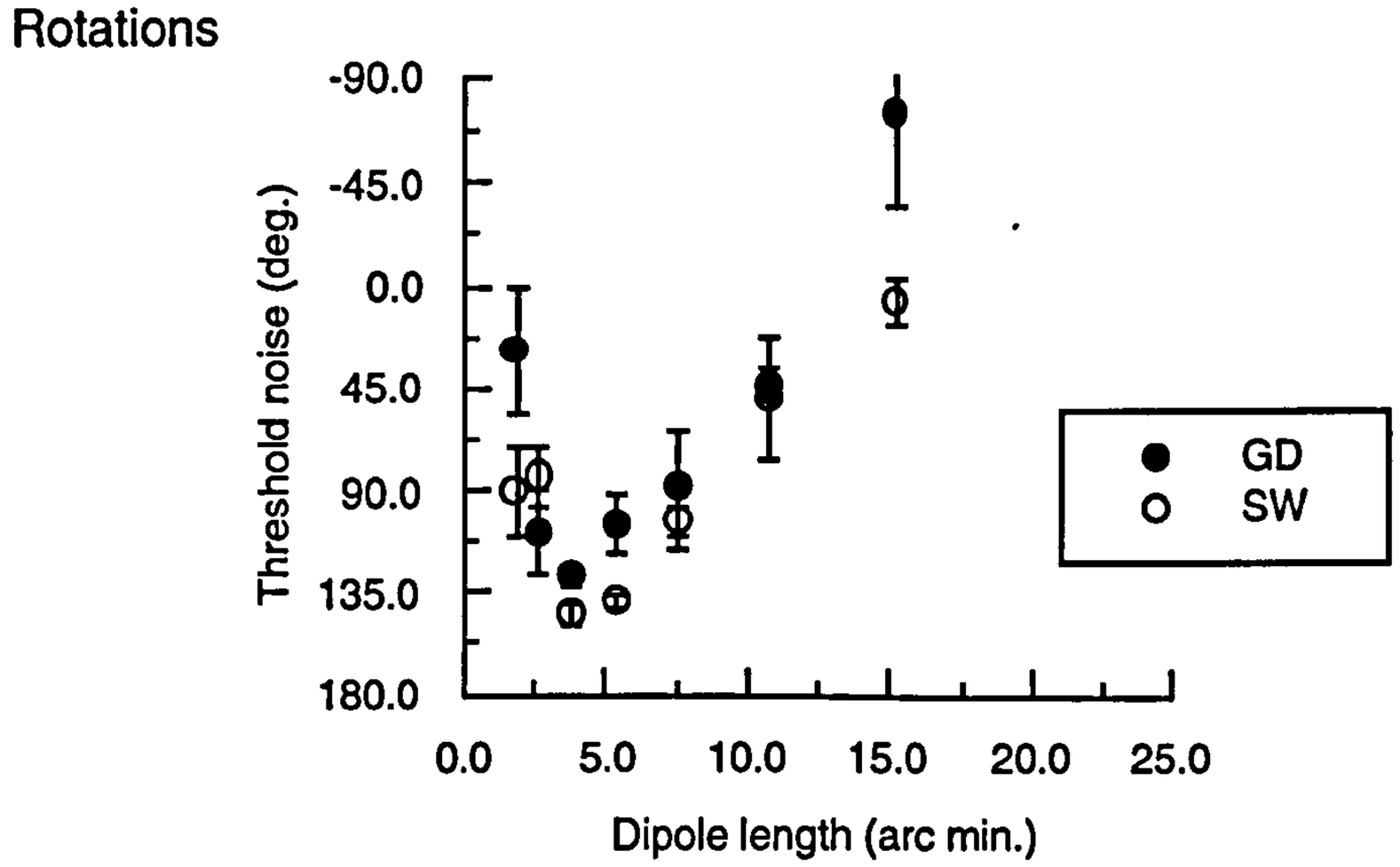


Figure 6.6. Experiment 6.2 results from three subjects. Graphs show the largest tolerable jitter of dipole orientation for a noise versus structure discrimination as a function of dipole length. Tolerable jitter is the width of the uniform orientation distribution, in degrees, that subjects could discriminate from a texture with purely random orientations.

6.5.2 Results

Results with two subjects for three transformation types are shown in Figure 6.6. The points plotted represent the largest tolerable jitter for the structure versus noise discrimination task. Negative and zero values, on the ordinate, indicate that the subject could not reliably perform the discrimination in the absence of orientation jitter. However, even though subjects cannot achieve 100% performance with no jitter present, the slope of the psychometric function, in the range tested, is still informative and so these values are presented. There do not appear to be any systematic differences between the data derived using different transformations: for all transformation types tested there are a small range of dipole separations at which the subject can withstand the most orientation jitter, typically at dipole lengths of 4-8 arc min. At this separation, subjects can reliably discriminate between noise and Glass patterns which are composed of dipoles which have the correct local orientation ±56°. The poor performance exhibited with textures composed of very short elements could be due to a number of factors. Even though there is no uncertainty as to the matching of the components of dot pairs, subjects are poorer at estimating the orientation of features with small aspect ratios (Westheimer, 1981). This is confirmed by data from Experiment 4.1 which showed that subjects are optimal at estimating the mean orientation of translational Glass patters at dipole separations of around 4-5 arc min. One complication, however, is in the use of the threshold orientation variance as a measure of performance. Because dots are sampled from a pixel array, variance might be expected to have an unpredictable effect on performance at very low separations. A final possible reason for reduced performance at small dipole separations, is that the spatial frequency of mechanisms sensitive to such dipoles is very high. Poor performance may be attributable to dipoles failing to exceed the contrast threshold of such filters.

The decrease in tolerable orientation variance as length increases is to be expected. As matching uncertainty increases the effect of orientation variability will be more disruptive.

6.5.3 Details of simulation

The two measures of pattern structure, field smoothness and template match (both described in the Experiment 6.1) were compared on the task described above. Simulations were run using identical stimuli to the psychophysical experiment; the same range of dipole length and the same transformations were simulated. However, the full version of the adaptive filtering model uses 7 spatial scales and 12 filter orientations. Even using

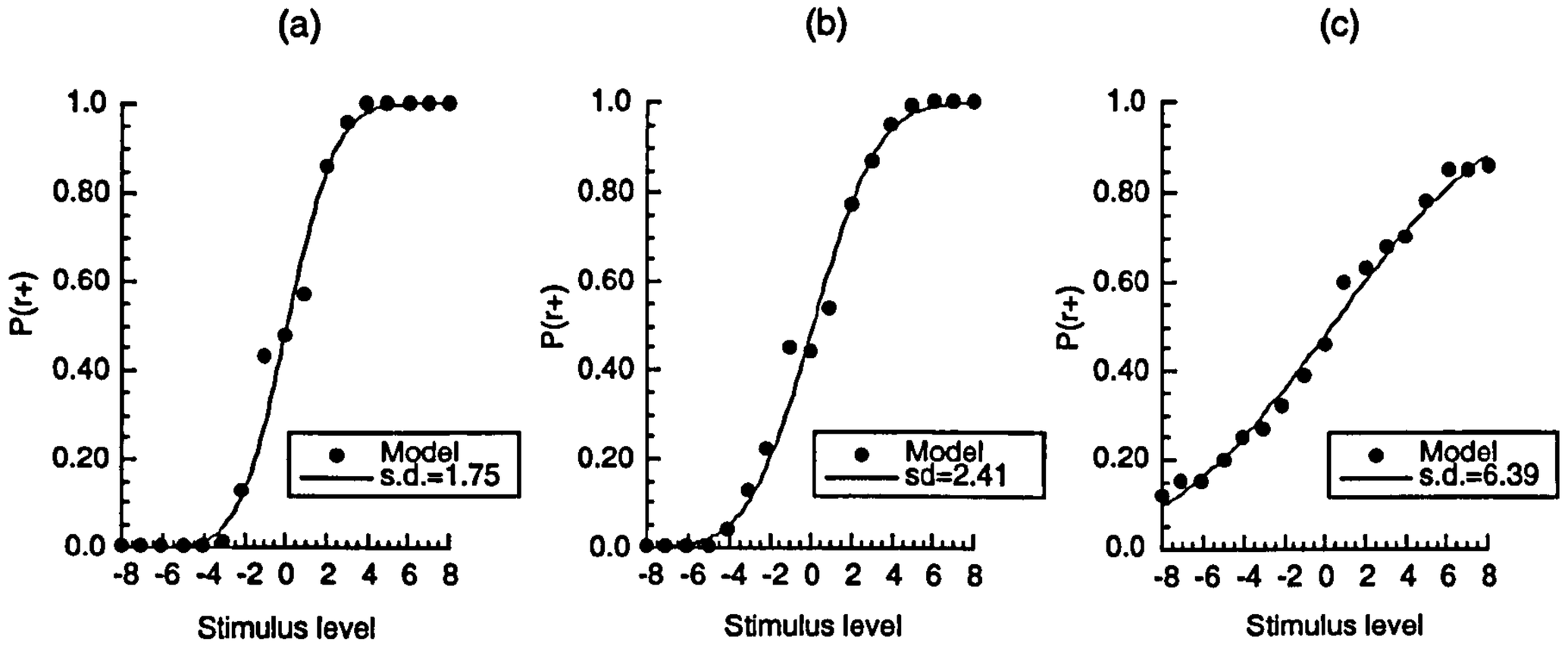
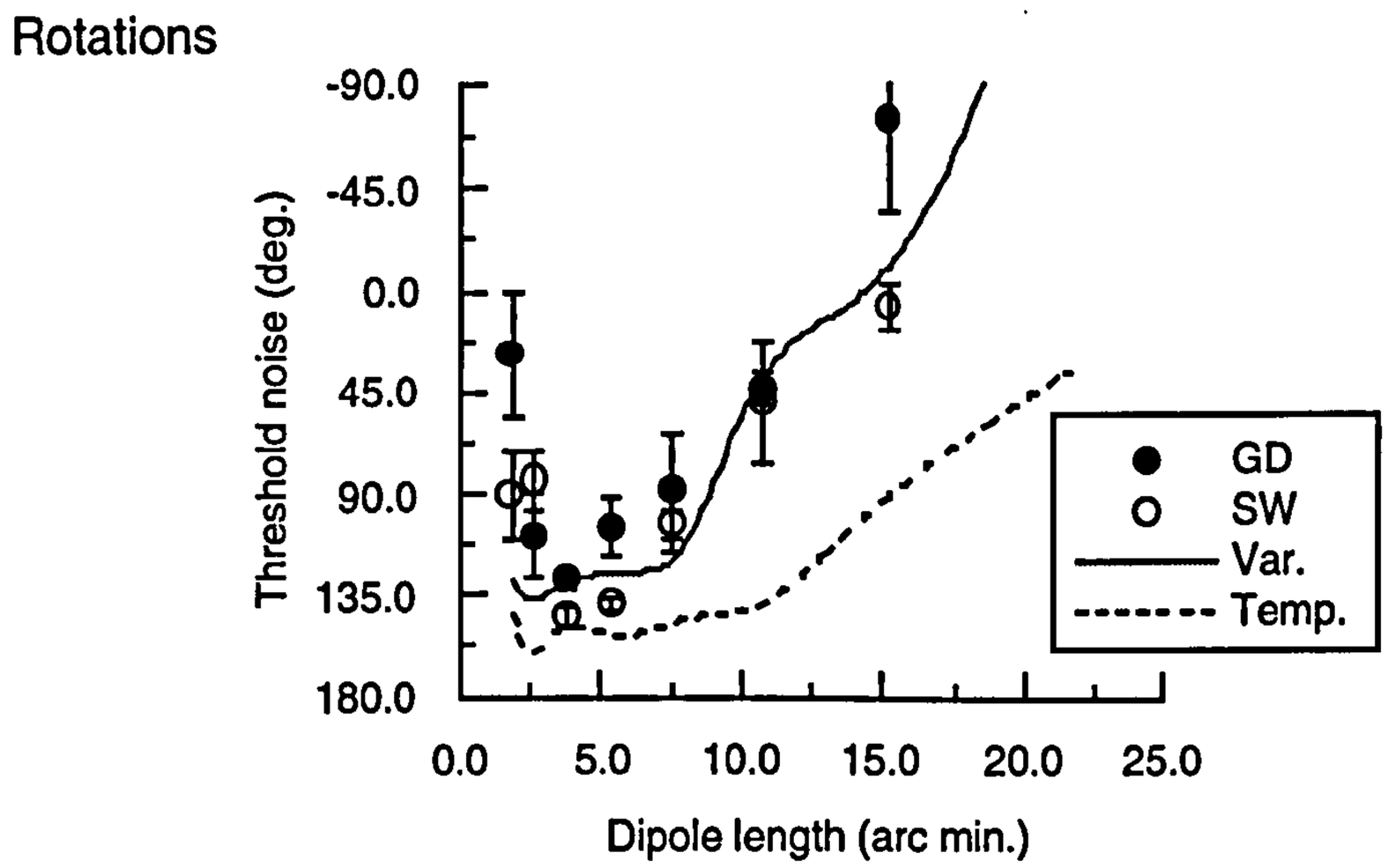


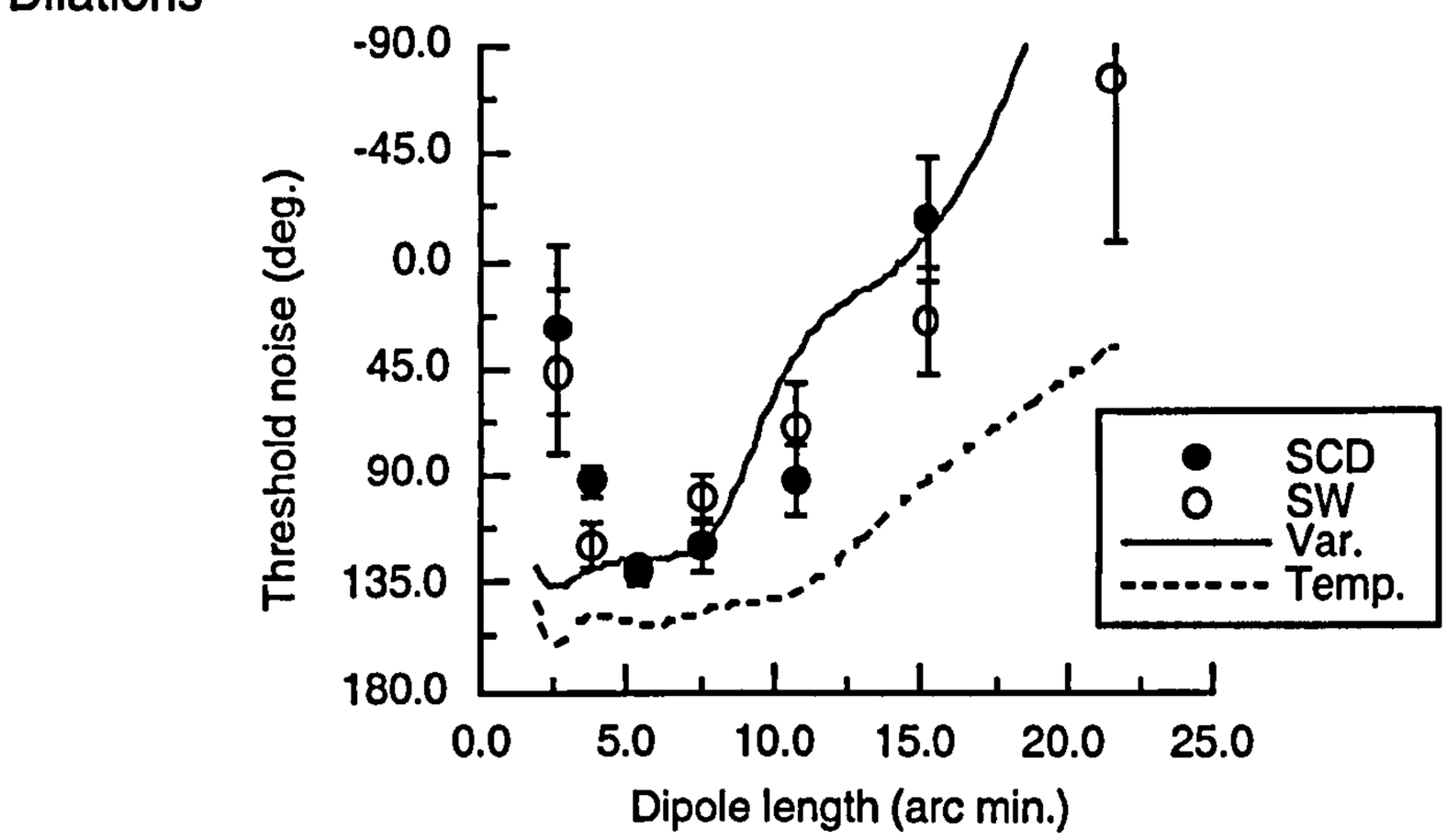
Figure 6.7. Typical psychometric functions from the simulation. These figures show the performance of the variance discriminator for dipole separations of 4.0, 8.0 and 16.0 arc min.

fast Fourier techniques it is not feasible to simulate 50 psychometric functions using 96 separate convolutions for every image. To avoid this problem a simplified version of the model was used. At the beginning of each new run, at each dipole length, the average optimal spatial scale was calculated over 16 Glass patterns and subsequent convolutions were only performed at *this scale*.

6.5.4 Simulation results and discussion

Psychometric functions were calculated at each dipole length condition by generating a set of 32 organised textures at each of the nine stimulus levels (−90° to +90° in steps of 22.5°) as used in the psychophysical experiment (a negative stimulus value indicated that the random pattern was presented in the first rather than the second interval). The variance and template-matched estimates were calculated for the pattern and these 32 values for the patterns were stored. A similar procedure was used to generate structure estimates for a reference/disorganised Glass pattern. The psychometric function could then be generated by sampling a pair of values from a file derived from cued stimuli and from one derived from the reference stimuli. The proportion of times one was greater than the other represents the probability that the model would discriminate a stimulus with a particular cue from the reference. Examples of psychometric functions derived in this way are given in Figure 6.7. Note that they are well fitted by the standard cumulative Gaussian model.





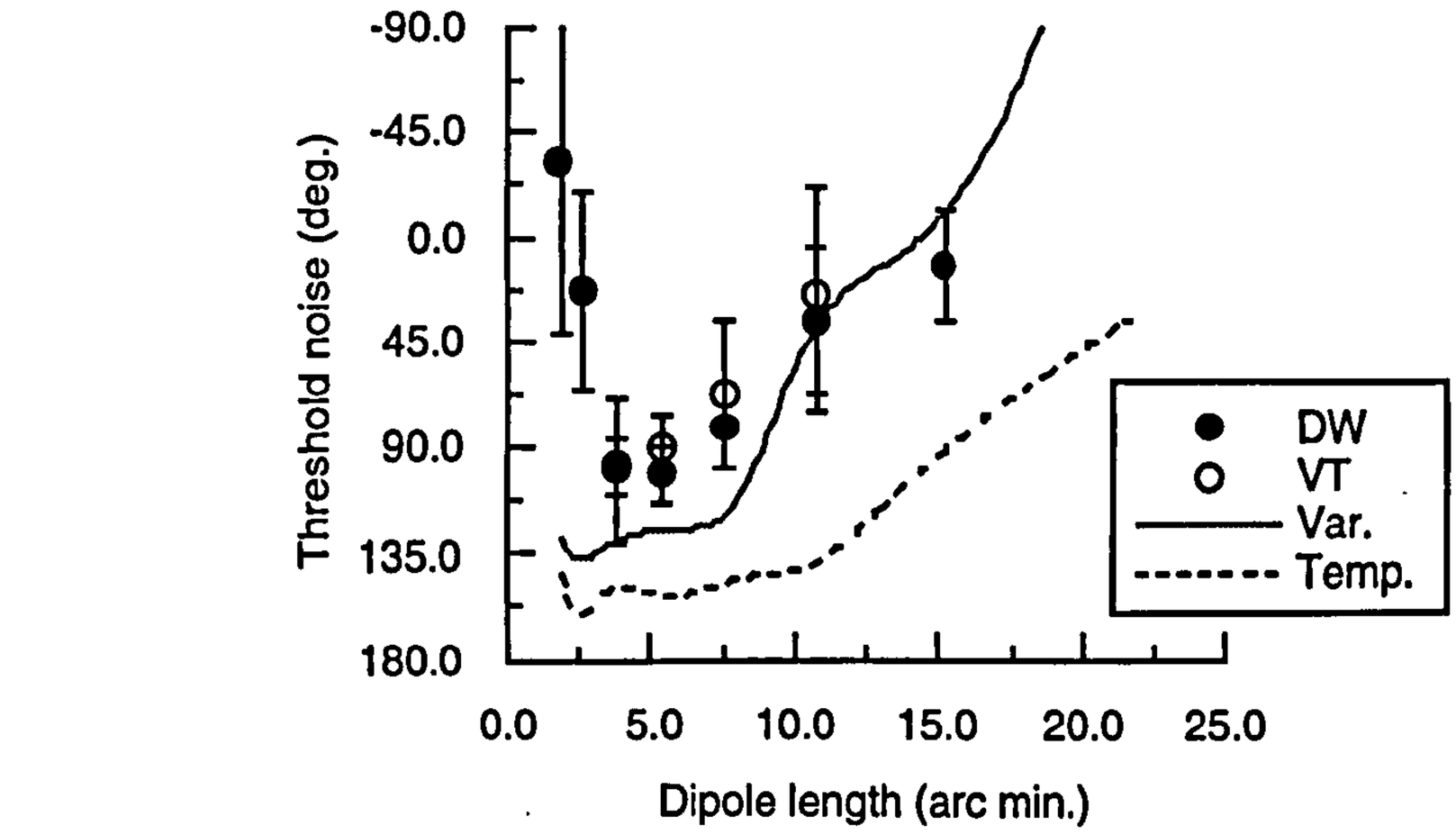


Figure 6.8. Simulation of Experiment 6.2 using the variance and template models. Note that the performance of both models deteriorates as a function of increasing, but not decreasing, dipole length.

$$P(r) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{r} \exp[\frac{-(e-\mu_e)^2}{2\sigma_e^2}]de$$

The standard deviation of the psychometric functions from the simulation are superimposed on human data in Figure 6.8. Firstly notice that the template model produces very high performance on this task. Discrimination at separations of around 2.5 to 10 arc min. was achieved when up to ±75° of variation was permitted. The best performance of the variance model was discrimination with approximately ±65° of variation, at separations of 2.5 to 6 arc min. The second point of interest is that the models show similar deterioration in performance to human subjects as dipole separations exceed around 8.0 arc min. No fitting is applied to the predictions, but the agreement between the variance model, in particular, and human data is good. The final observation from the simulation results is that neither of the models show any deterioration in performance at very short dipole lengths. In other words, human observers are unable to make use of the orientation information that is present in patterns containing very short dipoles. This suggests that subjects' performance is not due to pixelation effects, since both human subjects and models were given the same stimuli. Approximating a dipole with 1.5 cycles of a square wave, the mechanisms sensitive to a pattern with a dot separation of 4.0 arc min. are operating at about 23 cycles per degree. Such mechanisms are known to have much lower contrast sensitivity than mechanisms operating within the middle spatial frequency bands (e.g. using 6 Hz temporally modulated gratings, contrast sensitivity at 1 c.p.d. is ten times higher than for 20 c.p.d. stimuli; Robson, 1966). Given the long viewing distance (2.0 metres), and the limited spatial extent of the stimuli, it is possible that poor performance on this task is due to a failure of elements to consistently exceed contrast threshold of the mechanisms necessary to process them.

To summarise, this experiment suggests that the representation scheme proposed retains the global orientation structure of flow patterns, in the presence of noise, in a manner broadly consistent with human performance. However, the reduced performance of subjects when dealing with orientation structure confined to very high spatial frequency bands (> 23 cycles per degree) is not predicted by the model. This could be corrected by the addition of noise at high spatial frequencies.

6.6 Experiment 6.3: Detection of disruptions to flow in Glass patterns

In forming a smooth flow field, some averaging of orientation estimates over space must occur. At first sight it would seem easy to ensure that a flow field were always smooth by picking very large local regions of integration. This, however, would have two unfortunate side-effects. The first, which was considered in Section 6.5, concerns the smoothing out of structure in globally organised flow fields. The second is that any local discontinuities in the flow field would also be smoothed out. Such discontinuities are almost certainly interesting to the visual system as they can signal the occlusion of two surfaces or a surface edge (Hel Or and Zucker, 1989). The questions that this section addresses are: can the model proposed retain interesting discontinuities, and how does the model's performance compare to human performance at detecting disruptions to local flow? In order to establish how accurately discontinuities are retained in human vision, a psychophysical experiment is reported that examines disruptions to flow in Glass patterns.

Moraglia (1989) showed that detecting a horizontal line element within a circular flow field is easiest when the orientation of that element maximally disrupts local flow direction. Nothdurft (1992) (Experiment 3) used textures composed of short lines, containing single target lines which differed in orientation to the local orientation of the background field. The required orientation shift, for discrimination of the target from the background, increased as a function of the orientation variation of background elements. In accord with previous work (e.g. Nothdurft, 1985b), Nothdurft (1992) proposes that the detection of anomalous elements is not attributable to the use of texel alignment cues, but instead to the presence of local orientation contrast at the border.

The psychophysical procedure used in this section to investigate the detection of flow discontinuities is an adaptation of a paradigm described in Hel Or and Zucker (1989). They used Glass patterns, as examples of typical densely organised texture flows, which had small holes inserted in them. These holes were filled with unstructured dots. The threshold size of such embedded patches was determined for discrimination of these textures from ones which did not contain patches. The density of Glass patterns and the "path-length" of each texture element (the number of "overlays" used to generate the pattern) were systematically varied. Hel Or and Zucker (1989) found that the ability to detect anomalous patches was largely independent of the overall density of the pattern. Their claim was that it was primarily *path length* that determined performance. The

explanation offered, in terms of curvature constraints on matching, is considered below.

The psychophysical data presented in Hel Or and Zucker (1989) are only from one subject, and for this reason a replication of the original study was performed. The only major change was the use of an adaptive procedure to sample the cue.

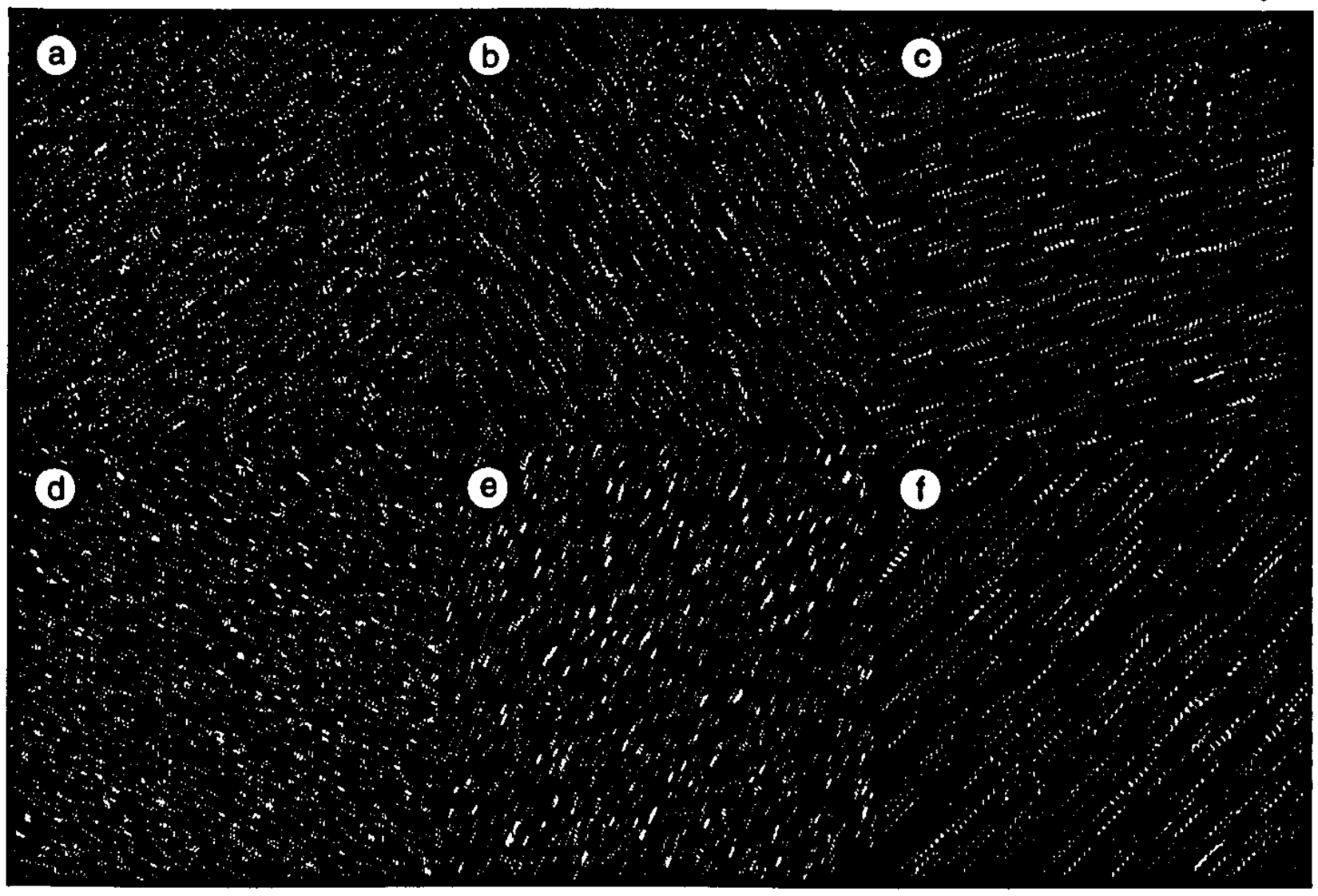


Figure 6.9. Examples of the stimuli used in Experiment 6.3. The three columns (from left to right) show textures with a path length of 2, 3 and 6 dots, respectively. Textures in the top row contain patches of unoriented noise with radii of 32 pixels, the maximum cue size tested in the experiment. The embedded patch in the bottom row has a radius of 16 pixels. Note that the embedded patch becomes progressively easier to detect as the path length increases.

6.6.1 Psychophysical procedure

The procedure used was similar to the one used in Hel Or and Zucker (1989), except that an adaptive method was used for data collection, making more efficient use of subjects. Unless stated otherwise, stimulus parameters were set to agree with Hel Or and Zucker (1989).

Subjects

Two subjects were used in this experiment. One (FJM) was naive to the purpose of the experiment and had normal vision. The other was the author (SCD) who had corrected-to-normal vision.

Apparatus

All stimuli were presented using the same equipment as described in the General Methods section of Chapter 4. Subjects were seated in a dimly lit room, 55 cm. away from the monitor.

Stimuli

The stimuli used were translational Glass patterns, created by randomly distributing texture elements, composed of collinear equally spaced dots, within a square field. As the number of dots making up the element (the *path length*) increases, the overall orientation of the pattern becomes progressively clearer (see Figure 6.9). Dots were composed of single pixels, subtending 2.0 arc min., and appeared white on a black background. The overall orientation of the patches was randomised from trial to trial. The separation of dots within each texture element was also randomised within the range 2-4 dot widths (4-8 arc min.). Patterns were presented within 256 by 256 pixel images which were 87 mm square. At the viewing distance of 58 cm, images subtended 8.53 degrees square.

Cued textures contained an anomalous embedded field of unoriented dots. This was created by allowing oriented texture elements to fall anywhere within the image except a circular region, with radius $r°$, creating Glass patterns with small holes. These holes were filled with non-oriented random dot fields, the density of which was identical to the background density of the Glass pattern (precluding the use of a local luminance cue to perform the task). Examples of the stimuli are shown in Figure 6.9.

In order to control for the effects of retinal eccentricity the centre of the anomalous patch was randomly located on an annulus around the centre of the pattern, with radius $r + 2.0°$.

Procedure

A two interval, two alternative forced choice task was used. This was to decide which of two sequentially presented Glass patterns contained an anomalous field. Stimuli were presented for 100 ms. The two stimuli were both preceded by a fixation marker which was presented in the centre of the screen for 500 ms. An additional 250 ms of delay followed the offset of the first pattern, making a total ISI of 750 ms. Subjects indicated their choice on the computer keyboard.

APE (Watt and Andrews, 1981), an adaptive method of constant stimuli, was used to determine the threshold patch size for discrimination. A range of patch radii from 0.0 to 32.0 pixels (64 arc min.), were adaptively sampled in steps of 4.0 pixels (8 arc min.). Runs of 64 trials were used with no interleaving. Dot densities of 2, 4, 6, 8 and 10% were used and the path lengths tested were 2, 3, 4, 6 and 8 dots.

6.6.2 Psychophysical results

The major difference between this and the Hel Or and Zucker (1989) study is the use of an adaptive procedure, rather than a method of constant stimuli, and subsequent measurement of thresholds at the 83% and not the 75% level. A sample of the psychometric functions from one subject is shown in Figure 6.10 along with fits from the standard cumulative Gaussian psychometric function.

Thresholds are presented for both subjects in Figure 6.11 as a function of path length. The patch size required for discrimination falls steadily as a function of path length. Density does not seem to affect the shape of this function and for this reason data are also presented averaged across density in Figure 6.11c.

In order to compare data from this experiment with those from the original study, thresholds for the 75% level were also calculated. Results from two subjects are shown in Figure 6.11d, along with data from Hel Or and Zucker (1989). The first set of plots shows the patch size required to perform at 75% correct as a function of path length and texture density. The agreement between the original study and this experiment seems reasonable, although the thresholds from the original study are considerably lower for short path lengths than the new data. Most importantly, however, both sets of data show that the minimum patch size for performing the task depends primarily on path length and not density.

There are two differences between these data and data from the original study that are of interest. Firstly, this experiment shows increasingly lower thresholds as path length increases, whereas the data from Hel Or and Zucker (1989) seem to asymptote around a path length of four. This is almost certainly due to the fact that an adaptive procedure was used in this experiment which allowed sampling of smaller patch radii than eight pixels, the smallest patch size tested in Hel Or and Zucker (1989). The second difference is that the data from Hel Or and Zucker demonstrate that the required size of patch to perform the task decreases most rapidly as path length changes from two to three.

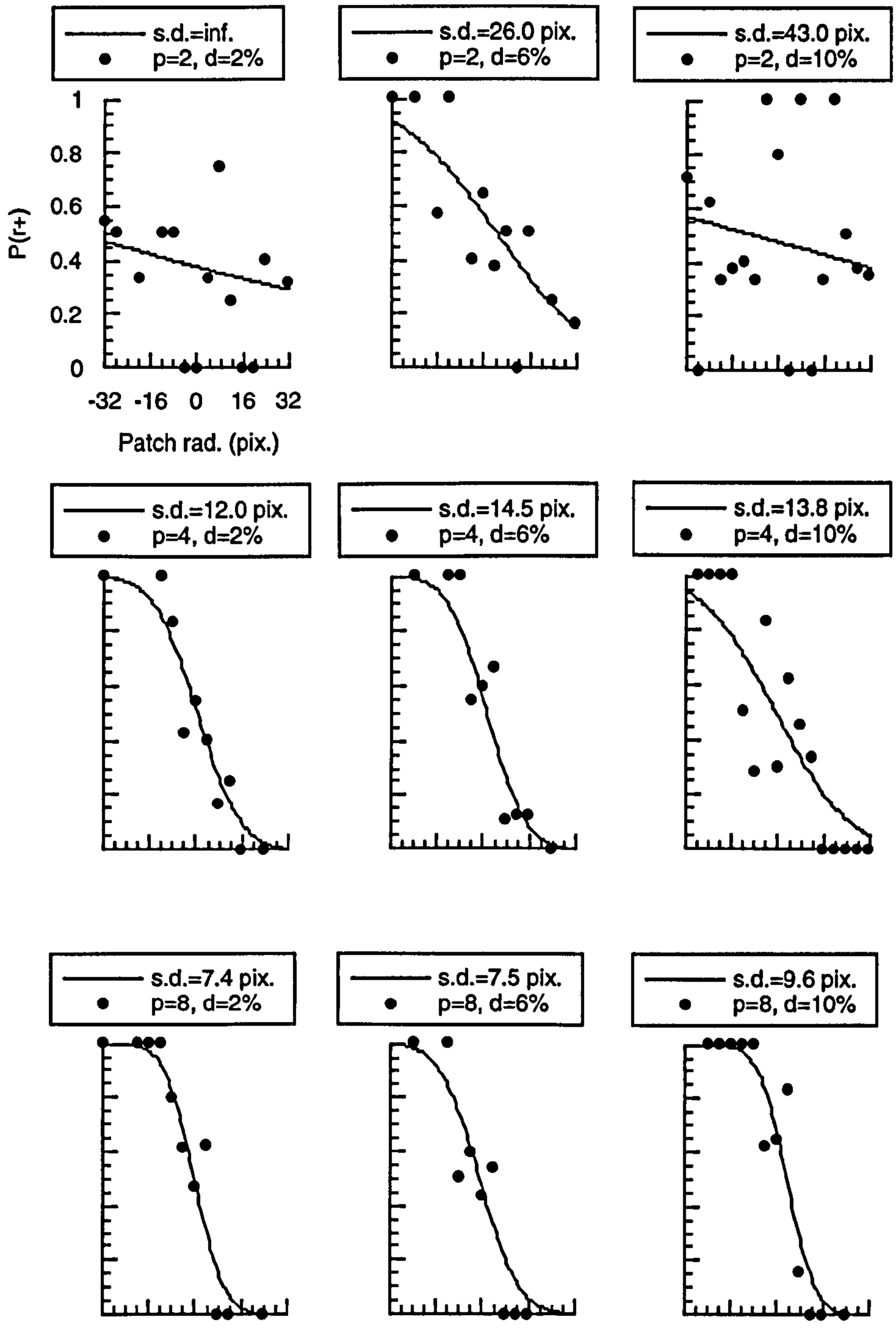


Figure 6.10. Typical psychometric functions for the anomalous patch detection task. Axes legends are shown in the top left-hand graph. In the legend boxes p refers to the patch length and d the flow density. The sign of the abscissa, the patch size, indicates whether the cued texture was presented in the first interval (negative) or the second (positive) interval. The ordinate shows the probability that the response was that the cued texture was in the first interval. Note that fits to the data shown, using the standard cumulative Gaussian psychometric function, provides a reasonable fit to most of the data.

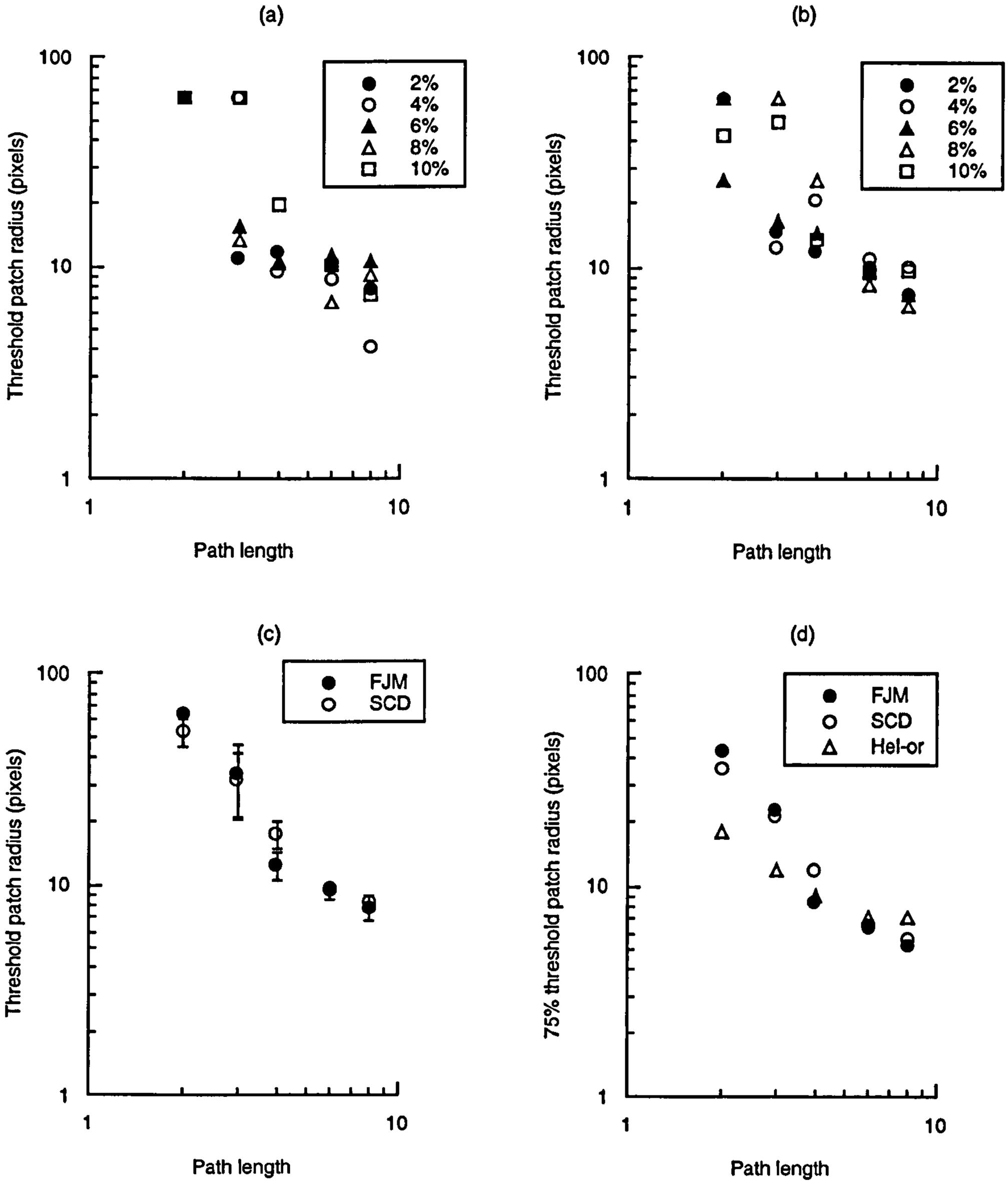


Figure 6.11. Data from Experiment 6.3. (Top row) Threshold patch radius for discrimination of textures with and without anomalous patches for two subjects (FJM and SCD respectively) as a function of path length. Note that data from all densities shows the same trend. (Bottom left) Data from (a) and (b) averaged across densities. Note the steady decrease in required patch size as path length increases. (Bottom right) Comparison of subjects detection of anomalous patches from this and Hel Or and Zucker (1989). Data are re-expressed as 75% thresholds and averaged over all density conditions.

However, this experiment shows a more steady increase in performance as a patch length increases. These data suggest that there is nothing special about the change in path length from two to three, a major component of the explanation of the data offered by Hel Or and Zucker and summarised below.

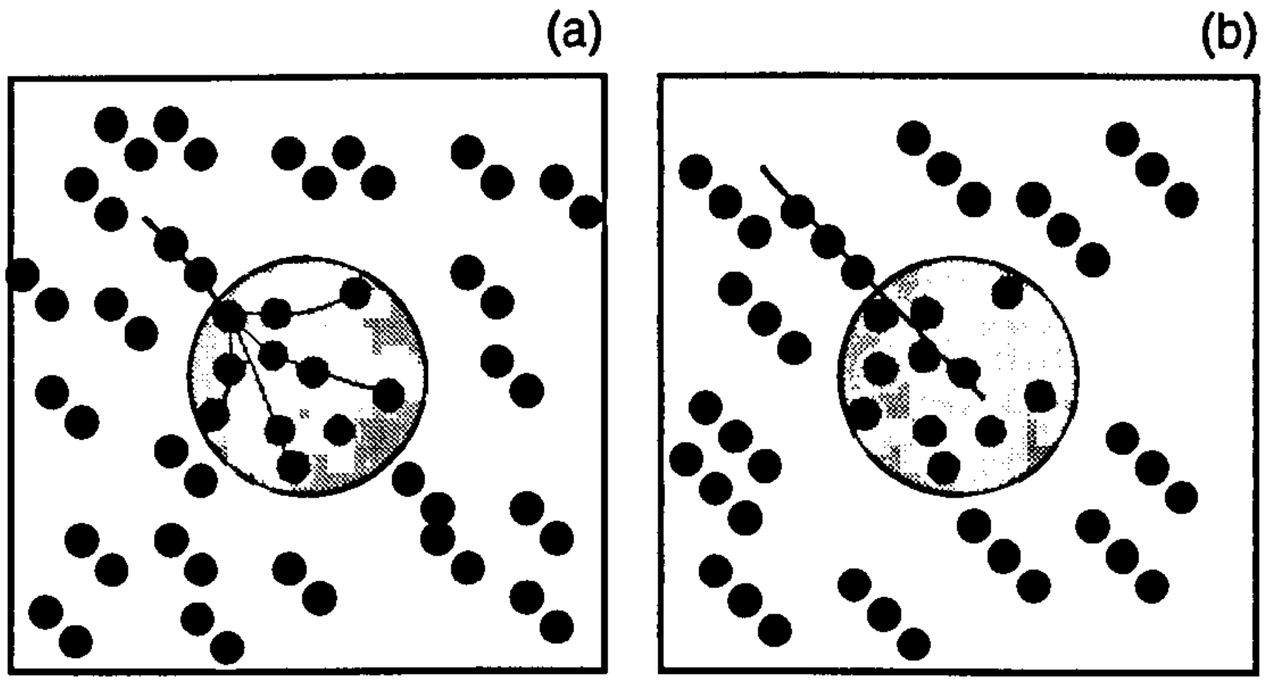


Figure 6.12. Hel Or and Zucker (1989) curvature hypothesis, adapted from Hel Or and Zucker (1989), Figure 4. (a) Shows a flow with path length two. There are many potential matches between elements near the patch and elements in the patch because paths are not constrained by curvature. (b) A flow with path length three. Curvature limits potential matches and increases the probability of detecting the patch.

6.6.3 Hel Or and Zucker (1989) interpretation

In order to propagate and smooth the flow field given by the output of a set of oriented receptive fields, Hel Or and Zucker propose that an interactive, cooperative procedure is used. This refinement occurs through inter-columnar interaction in V1 (Parent and Zucker, 1989) and operates by modifying the firing rate of cells within a neighbourhood according to their mutual consistency with respect to orientation. This amounts to propagating curves within the image. However, at the border of a texture patch orientation information is ambiguous and, Hel Or and Zucker claim, curvature information is necessary to stop the smoothing out of the discontinuity.

Figure 6.12 illustrates this explanation. The left hand part of the figure shows a flow field with a path length of two. At the patch border a wide range of possible matches are possible, because matches are unconstrained with respect to curvature. There will be enough random matches that match the background orientation to ensure that smooth patches will be smoothed out. For the case of a flow with path length three, illustrated in Figure 6.12b, matches at the border are constrained to have zero curvature and so many fewer matches are possible. Consequently, patches are less likely to be smoothed out.

There are a number of problems with this explanation. Firstly the result that it is hard to detect anomalous patches in textures with a path length of two does not necessarily mean that recourse need be made to an explanation in terms of curvature. It is quite possible that a extra dots in the path merely increase the differential output of a filter at a

particular scale or orientation and it is this increase in energy, relative to the output over the anomalous patch that improves performance. Since the curvature model has not been implemented its adequacy has not been tested quantitatively. An alternative explanation using orientation reliability is given in the next section.

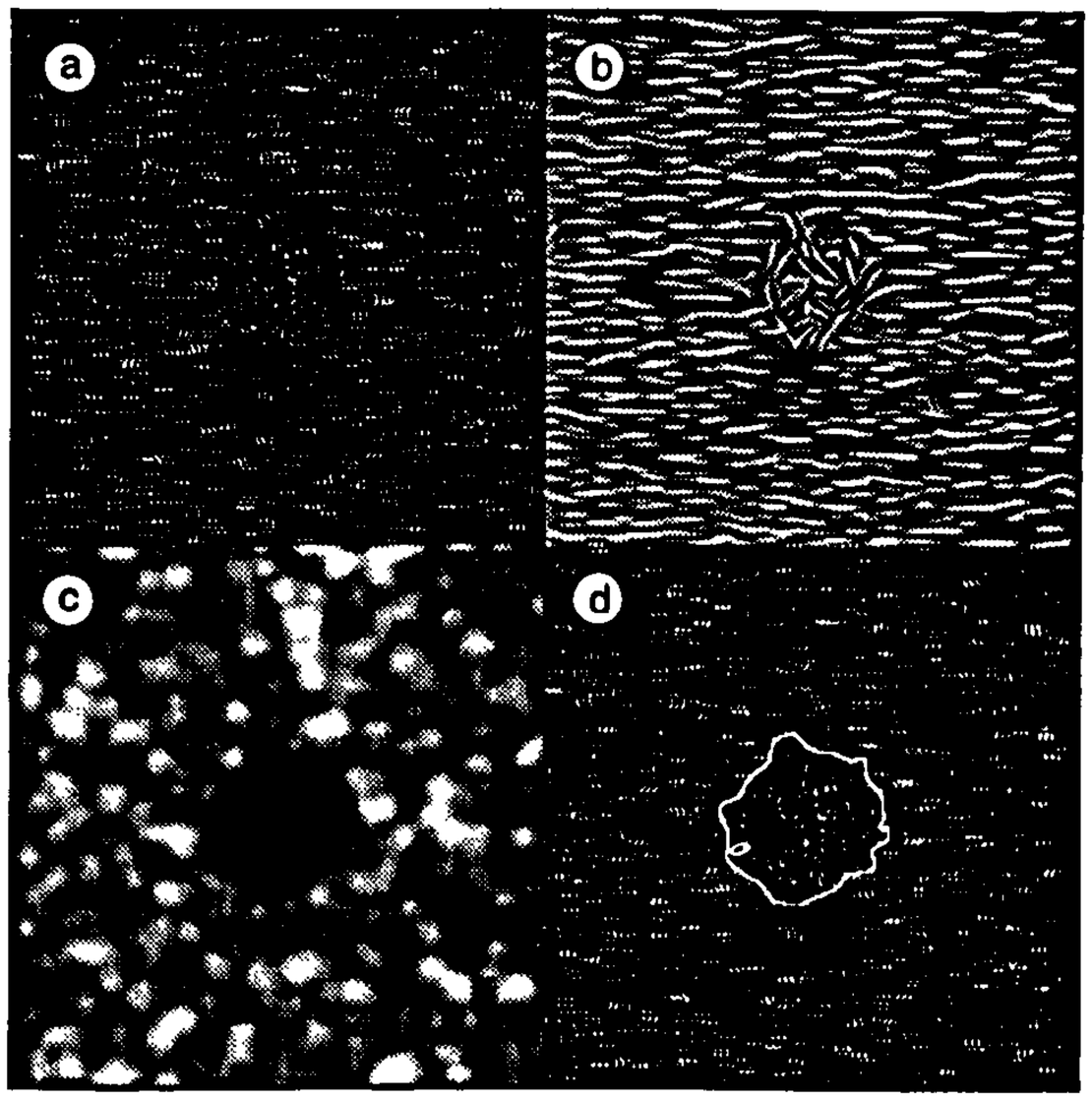


Figure 6.13. The boundary detection model. (a) A typical texture from Experiment 6.2 (b) The outputs from the scale-orientation combination rule (c-d) The energy at 0° and 90° respectively, at that spatial scale. (e) The ratio of (c):(d). Note the region of high energy around the embedded patch. By thresholding this image and examining the zero-crossings the boundaries of the patch may be located (as shown superimposed on the original image in (f))

6.6.4 An alternative interpretation

In this experiment, anomalous patches were marked out from the background flow by the fact that they were un-oriented. In terms of the adaptive filtering model, we would expect such a patch to be signalled by a region of high local orientation variance. The operation of the model on a stimulus pattern is illustrated in Figure 6.13. Figure 6.13c shows that the anomalous patch is indicated clearly by a region of low anisotropy. This suggests an alternative strategy for performing the discrimination in this experiment: estimate anisotropy for both images and select the image which contains the largest region of anisotropy. A number of simple additions have to be made to the adaptive filtering model to decide if the image contains a patch. Specifically we need to find a region of low anisotropy and measure its size.

- Adaptively filter at multiple scales
- Select the spatial scale which minimises orientation variance
- Threshold the anisotropy image above and below one grey-level standard deviation
- Characterise the resulting set of blobs using the image description scheme and locate the blob with the largest (negative) mass.

From here the boundary of the patch can be determined by simply using the boundary of the blob (Figure 6.13d). A 2AFC can be performed by selecting the image with the largest mass blob in the anisotropy map. This was the *boundary detection* model used to simulate the experimental task.

6.6.5 Details of simulation

A simplified version of the model was again used, due to processing time restrictions. The average optimal spatial scale was calculated over 32 Glass patterns at the beginning of a run, and subsequent convolutions were only performed at *this scale*[1].

The same five dot densities (2%, 4%, 6%, 8% and 10%) and path lengths (2, 3, 4, 6 and 8 dots) were used as in the psychophysical experiments, giving a total of 25 psychometric functions to generate. For each run, 32 stimuli were generated at each of nine different patch radii from 0.0 to 32.0 pixels. The mass of the largest blob in the anisotropy map was derived for each stimulus and was recorded in a file. Having collected predictions from the model at each of the cue levels, a psychometric function was calculated by sampling from each stimulus level and calculating the probability of discriminating blob masses from blob masses derived from the reference images.

One variation on the boundary detection model was also simulated. The subjects' poor performance with Glass patterns containing short paths could be because of two reasons. Firstly because they failed to locate any perceptually anomalous regions of the cued image (Hel-Or and Zucker's assumed reason). Or secondly because, even if they knew exactly where the patch might be, small patches of Glass patterns are hard to tell from noise (Glass, 1969). It would be interesting to know how uncertainty in the patch's position effected performance. To investigate this the same model was used in another simulation, except that it received only a 128 pixel square patch of the 256 pixel square

[1]Even using this simplification, the simulation described involved approximately 92000 convolutions which took 15 days on a Hewlett Packard series 700 Apollo work station.

image which, if the pattern was cued, would contain the anomalous patch. This is the low spatial uncertainty condition.

Finally the template and local variance model were also run on the patterns. Disruptions to local flow may be signalled as changes in the global measure.

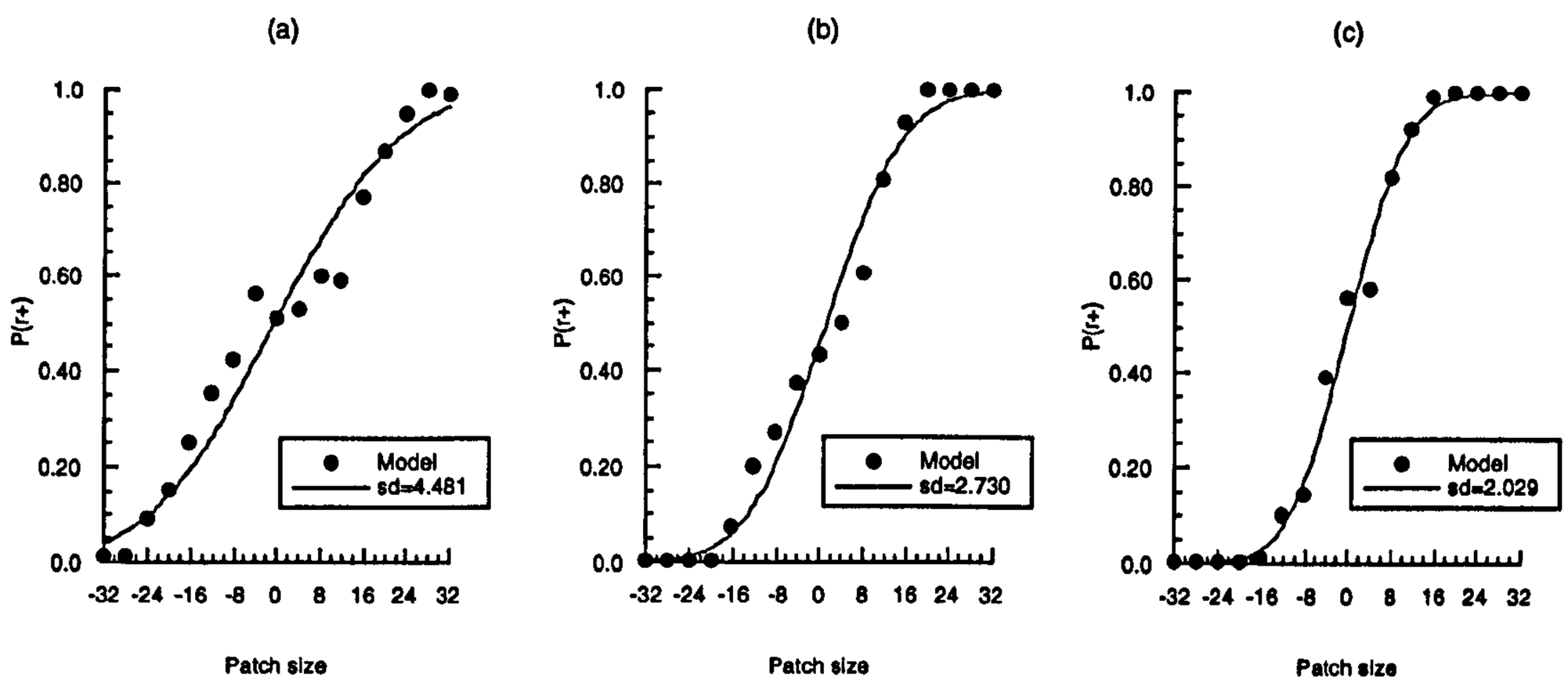


Figure 6.14. Typical psychometric functions from the simulation. This figure shows the performance of the discriminator using mean local orientation variance for textures with densities of 6% and path lengths of (a) 2, (b) 4 and (c) 8 dots.

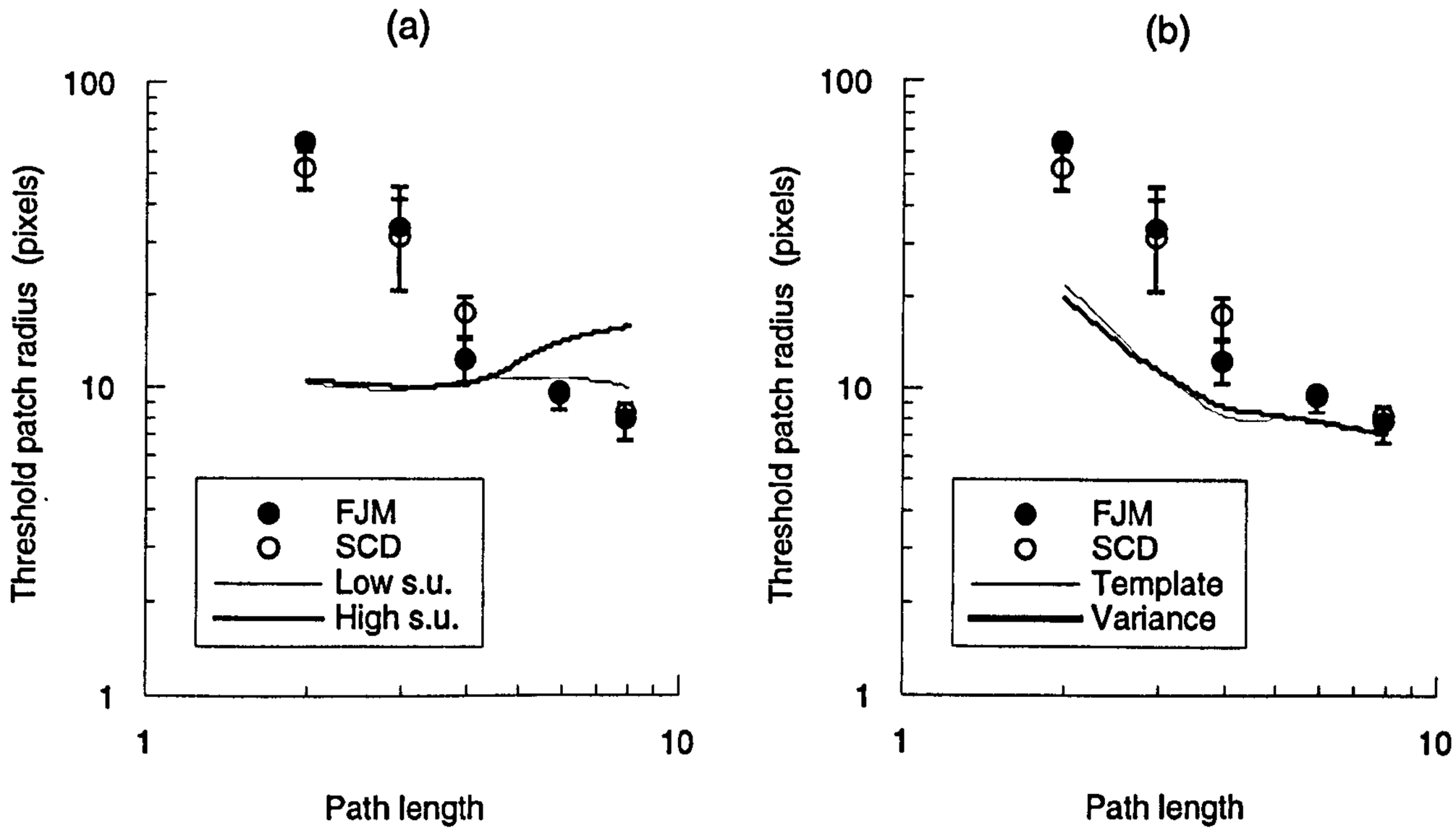


Figure 6.15. Results from the simulations of the task from Experiment 6.3. (a) Predictions from the boundary detection model. (b) Predictions form the variance and template matching models. Only the variance and template models produce performance consistent with human subjects.

6.6.6 Simulation results

The standard deviations of simulated psychometric functions, from the two variations on the boundary detection model, are shown superimposed on subjects' data in Figure 6.15a. Considering firstly the data from the low spatial uncertainty condition it is apparent that path length has little effect on determining which pattern contains a Glass pattern if the approximate location of the path is known. The threshold patch radius achieved by the boundary detection model (s.d. $\approx$ 12.0 pixels) compares well with the best performance of the subjects at long path lengths.

However the local boundary model fails completely when the entire pattern is presented to the model. The possibility that this was due to insufficiently dense sampling of local variance was investigated by re-running the simulation with different sampling densities. This made no appreciable difference to the predictions of the model. Given the complexity of the boundary detection model there are a large number of parameters which could be responsible for its failure.

Figure 6.15b shows predictions from the variance and template model for this task. The fit to human data is remarkably good given that each point represents the standard deviation of an entire psychometric function and that no fitting has been applied.

The success of the variance model strongly suggests that this task may not require any kind of boundary detection at all. On inspection, boundaries can be detected within the image, but the boundary model simulation suggests that this may not be as reliable a cue to discrimination, as simply estimating the overall smoothness of the texture. In order to preclude the use of simple global statistics in performing segmentation tasks, some authors (e.g. Nothdurft, 1985b; Landy and Bergen, 1991) use patch *shape* discrimination tasks, and a thorough comparison of these models for texture segmentation would certainly have to use data from studies such as these.

These simulation results are in broad agreement with those of Nothdurft (1992), since the variance model would also predict that background orientation variation will critically affect detection of anomalous elements. Nothdurft suggests that disruption to flow is not detected by breaks in the co-aligned texture elements but in local orientation contrast at the anomalous region. Using the pre-processing stage proposed in this thesis, breaks in local texel alignment would be coded by a failure to coalesce into single elements at coarser spatial scales (see Chapter 2, Section 2.7.4). Pilot trials demonstrated that this was not a reliable cue to the presence of an anomalous patch. Instead the proposed (variance)

model codes the sum of local orientation variance. This measure co-varies with the degree of local orientation contrast.

6.7 Conclusions

This chapter has used the adaptive filter model combined with a measure of mean local orientation variance, and a measure of the degree of local orientation matching to a template to simulate three psychophysical tasks. This was specifically intended to test the spatial aspects of the model by considering how globally organised patterns are coded. It has been shown that when these measures are treated as discriminators they approach and exceed human levels performance on three different psychophysical tasks, involving the perception of structure in Glass patterns.

Data from Maloney *et al.* (1987) (Experiment 6.1) demonstrated that whilst a measure of local orientation variance could predict the discriminability of Glass patterns in the presence of extra, noise dots there had to be sufficient texels present for the measure to be of any use. At low texel densities it was suggested that subjects make use of more detailed knowledge about the underlying transform.

The Hel Or and Zucker (1989) task is closely related to a variety of texture segmentation tasks. The success of the orientation variance model suggests that this task does not require the explicit coding of the *position* of texture boundaries, and that instead local orientation statistics suffice. This is illustrative of the problem of presenting subjects with segmentation tasks - the information they can use is often insufficiently constrained.

7 | THE EFFECT OF LOCAL CONTRAST ON GROUPING

7.1 INTRODUCTION

The spatial filters used in most models of early visual processing are *linear*; they transform the image intensity at a point into a weighted linear combination of neighbouring values. The view that the visual system decomposes the complex two-dimensional signal from an image into the *linear* sum of a number of terms in the frequency domain is convenient for two reasons. Firstly, the assumption of linearity allows the study of the visual system psychophysically through the presentation of patterns which only contain a small number of those frequency terms (*sine wave grating* studies e.g. Campbell and Robson, 1968). Secondly, it allows for computationally efficient simulations of visual processing using Fourier techniques - efficient methods for moving between spatial and frequency based representations.

It has been proposed that such filters also provide a natural mechanism for visual *grouping* (e.g. Watt, 1988). If two visual features both fall in the excitatory region which are below mean luminance, they will not produce an elevated response. Since it is the response of these filters that is proposed as a mechanism for grouping spatially distinct features, this suggests a problem. It should be possible to construct stimuli for which grouping will be eradicated when the contrast of the components is altered in such a way as to destroy the structured output of filters. If the breakdown of grouping predicted by a purely linear system does not correspond to human performance then a reassessment of the model would be required.

Two types of such contrast phenomena have been widely used to criticise models of visual grouping based on spatial filtering.

The first phenomenon is the perceived grouping of "balanced" features which (supposedly) do not contain the low spatial frequency information that filtering models are sensitive to. Stevens and Brookes (1987) showed that structure is correctly perceived in

Glass patterns composed of balanced energy features (each feature is made up of a bright centre surrounded by dark ring). They claim that whilst we can perceive structure in balanced energy patterns, models based on spatial summation will "smooth out" individual features and therefore cannot explain the perceived grouping. It will be demonstrated, in this chapter, that the adaptive filtering model can correctly predict perceived structure.

The second phenomenon is, that contrary to the prediction of filtering models, features of opposite contrast polarity will group under certain circumstances. If the correlated elements comprising a Glass pattern are of opposite contrast polarity then, it is claimed, orientation structure cannot be perceived (Glass and Switkes, 1976; Prazdny, 1986b; Zucker and Davis, 1988; Zucker *et al.*, 1983). However, Kovacs and Julesz (1992) have shown that, under certain conditions, rather than being destroyed, perceived flow is rotated through 90°. They present psychophysical data from an experiment which systematically varied the background luminance of Glass patterns as a function of the mean luminance of the dipoles. The adaptive filtering model will be used to generate predictions of human performance on this task.

More complex stimuli have been designed to "fool" filtering models. Prazdny (1986a) points out that the *correct* flow direction is perceived in a texture composed of two "interleaved" Glass patterns, each with opposite contrast polarity. Elements of these patterns are composed of a sequence of dots with positive, negative, positive, and negative contrast polarities (e.g. Figure 7.8). Linear filtering of these patterns does not produce the correct structure at any spatial scale, and Prazdny (1986a) takes this as strong evidence against models employing such mechanisms. In this chapter it is argued that an early non-linearity in the transduction of image intensities can adequately explain the perception of structure in these patterns.

7.2 Spatially high-pass textures

Many proposed filtering models for extracting tokens from images use mechanisms which are only sensitive to low spatial frequencies. Figure 7.1 shows (a) a typical Glass pattern, and (b) its power spectrum, with boundaries of the sensitivity of a typical horizontal DoG filter superimposed. The latter image is on polar axis: the distance from the centre indicates scale (0-128 cycles per image), and the orientation of a point from the centre, offset by 90°, indicates orientation. Brightness indicates power at that scale/orientation.

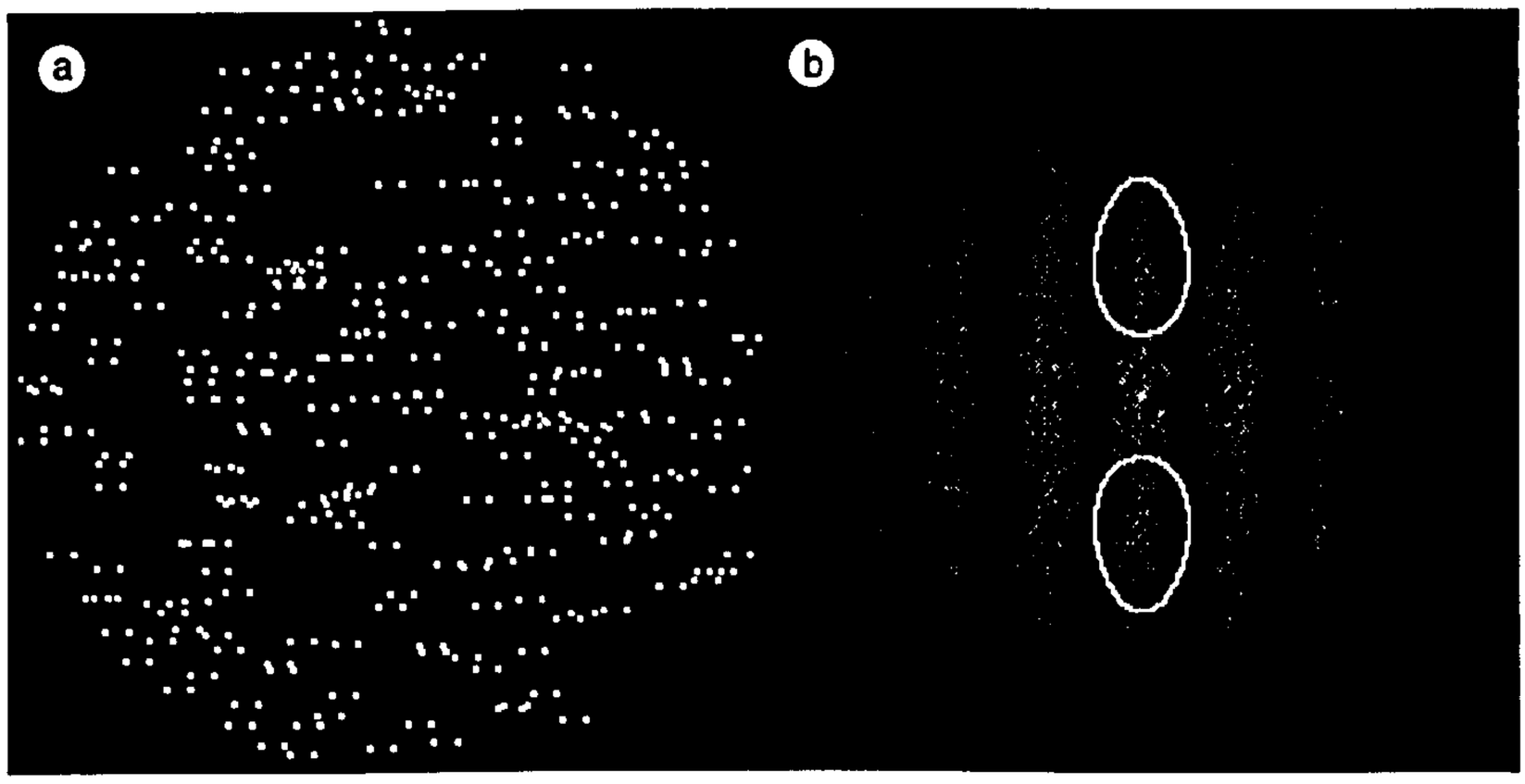


Figure 7.1. (a) Horizontal Glass pattern with dipole length of 8 pixels. (b) Fourier power spectrum of (a). Superimposed boundaries show the approximate limit of the sensivity of a DoG with peak sensitivity of 41 cycles per image. Note that the boundaries capture high energy regions of the spectrum, and that horizontal energy (the vertical column centred on the middle of the pattern) is spread over a large range of scales.

The vertical stripes of high energy in Figure 7.1b are due to the dot spacing of the dipoles.[1] Notice that the power spectrum contains horizontal energy over a wide range of scales, and that that energy extends over the origin. This indicates that low spatial frequency information in this pattern will signal the presence of horizontal structure in the pattern. This could be detected using a coarse scale, low-pass filter (e.g. a Gaussian).

An apparently simple way of testing such a class of model is to remove all low spatial frequency information from the image by high-pass filtering it. A simple approximation to the high-pass filtered version of a dot is the balanced feature (Carlson *et al.*, 1980). This consists of an inner area with luminance greater than the background, surrounded by a region with luminance less than the background. The values of the inner and outer region luminances are selected so that overall average luminance of the feature is the same as the background. Structure may be perceived in patterns composed of these high-pass features as has been demonstrated notably with "Gestalt" grouping (Janez, 1984; Palmer, 1992) and visual illusions (Carlson *et al.*, 1980, but see Garcia-Perez, 1991) and such work has been widely cited as evidence against low-pass filtering as a mechanism for visual grouping (e.g. Beck *et al.*, 1987; Reed and Wechsler, 1990; Palmer, 1992).

Prazdny (1986b), however, generated Glass pattern composed of small Laplacian rings

[1] If we approximate the bright-dark-bright luminance profile of a dipole with 1.5 cycles of a square-wave grating, then we would expect that as the scale of analysis of this pattern changed, the energy of this feature would vary periodically.

and claimed that, even at 1 second exposure duration, subjects attempting to distinguish between transformation types approached chance levels of performance when background luminance approached the same value as the mean of the Laplacian features. This is entirely contrary to predictions from other grouping studies using balanced features and may be interpreted as evidence for a grouping process which involves a degree of spatial summation. It is hard for models based on matching of symbolic tokens (e.g. Stevens, 1978) to accommodate such results because all features, balanced or otherwise, are identical (as long as they are visible) and should be strongly matched.

Stevens and Brookes (1987) found a result contradictory to Prazdny (1986b) using patterns composed of small balanced dots. Subjects were reported as being able to perceive correctly structure in these patterns at exposure durations of as little as 200 ms. Stevens and Brookes claim that the discrepancy between this and Prazdny's result is due to the fact that the features in their patterns were scaled by eccentricity. In order to perceive structure in globally organised Glass patterns it is known that one must be able to see a sufficiently large region of the pattern (Glass, 1969) and it is possible that, due to the effects of retinal eccentricity, the outer portions of Prazdny's fairly large (8.25° wide) patterns were not visible to subjects. Hence they could not discriminate between globally organised patterns. This interpretation ignores the fact that subjects in Prazdny's study had long enough to use eye movements. Also Stevens and Brookes rely entirely on demonstrations and present no quantitative study of the strength of perceived structure in these patterns. These objections aside, the proposition that structure can be perceived in these patterns which are "devoid of low spatial frequencies" (Stevens and Brookes, 1987), will be considered.

Stevens and Brookes (1987) claim that since balanced dots provide negligible input to a simple cell, whose excitatory receptive field they lie within, the addition of an early non-linearity in intensity transduction followed by stimulation of a *very* long receptive field (aspect ratio greater than 1:20), is necessary. Figure 7.2a shows a typical balanced pattern, and Figure 7.2b its convolution with a $\nabla^2 G$ filter. The $\nabla^2 G$ produces poorly oriented blobs but does seem to reflect the circular structure. Figures 7.2c-d show the output of the adaptive filtering model which, in contradiction to the Stevens and Brookes's prediction, correctly derives the flow field. Figure 7.3 explains why the models can produce this structure. Whilst low spatial frequencies are not present in the image, indicated by the hole round the origin of the image, there are abundant higher spatial scales at which

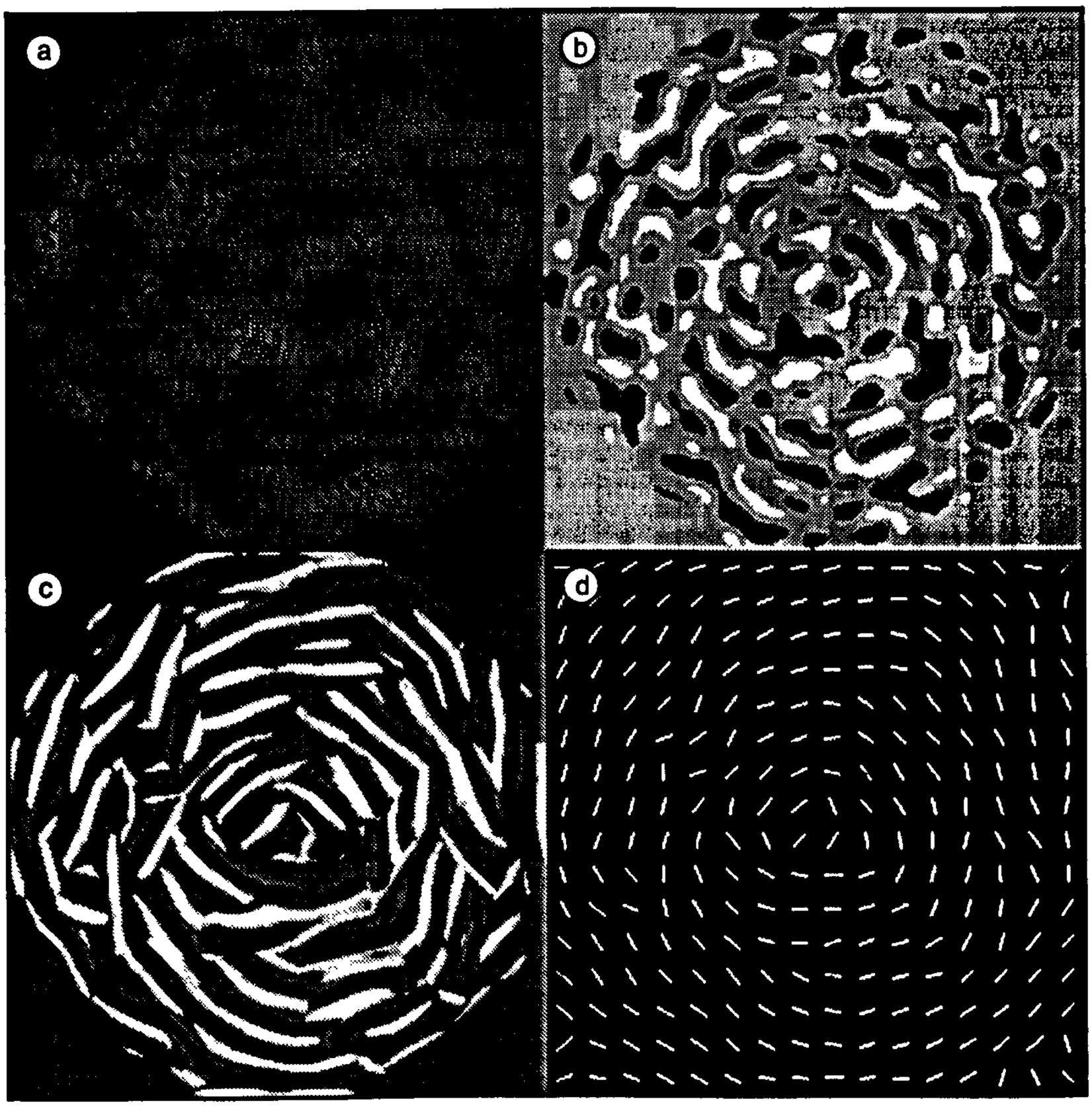


Figure 7.2. (a) A balanced Glass pattern (contrast enhanced for the purpose of reproduction). (b) $\nabla^2 G$ filtering poorly enhances orientation structure but output of the adaptive filtering model, (c), does much better. Note highly oriented features and derived flow field (d).

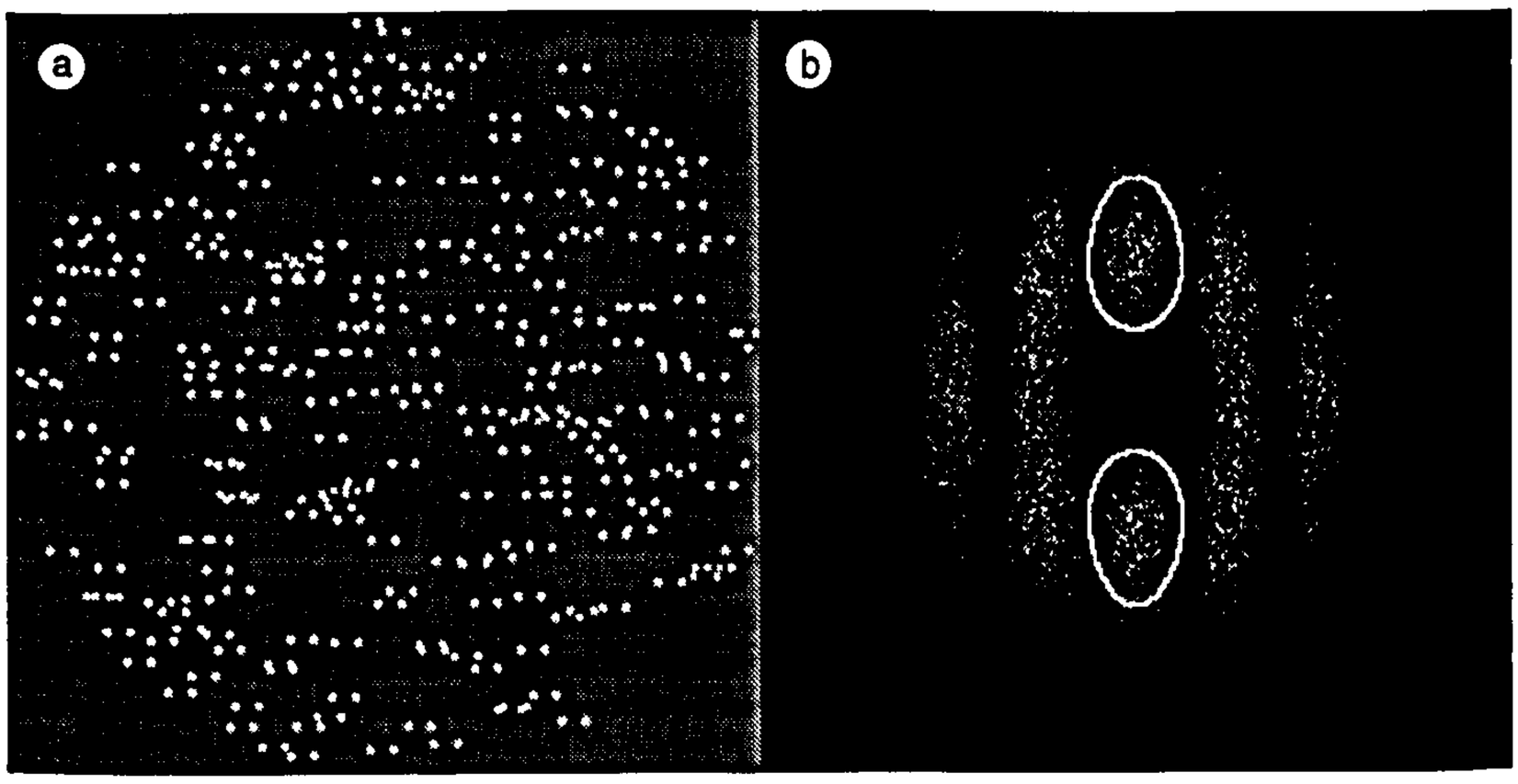


Figure 7.3. (a) High-pass filtered horizontal Glass pattern (Figure 7.1a). (b) Fourier power spectrum of (a). Superimposed boundaries show the approximate limit of the sensivity of a DoG with peak sensitivity of 41 cycles per image. Note that the boundaries capture high energy regions of the spectrum.

structure exists. The intersection of the DoG sensitivity boundary with regions of high energy indicates that such structure is readily detectable with an appropriately oriented feature.

While balanced features specifically disrupt the output of low-pass filtering mechanisms, models using band-pass filtering have no problem deriving local structure from Glass patterns composed of such features. We conclude that human perception of structure in these patterns offers no evidence against models of visual grouping based on band-pass (oriented or isotropic) filters. Prazdny's (1986b) finding that subjects could not see structure in these patterns seems likely to be due to the artifacts pointed out by Stevens and Brookes.

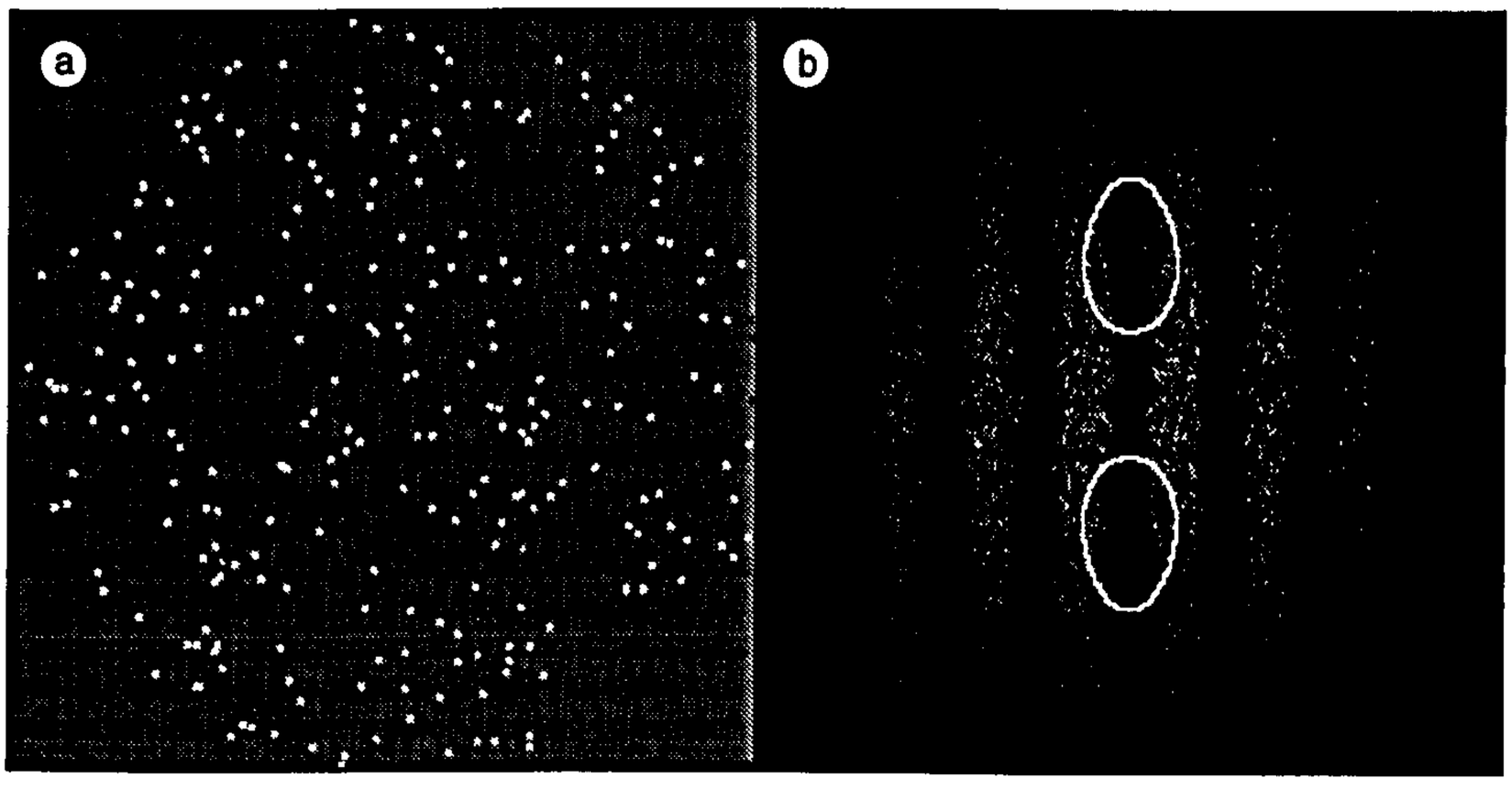


Figure 7.4. Horizontal Glass pattern, with dipole length of 8 pixels, composed of opposite contrast dots. (b) Fourier power spectrum of (a). Superimposed boundaries show the approximate limit of the sensivity of a DoG with peak sensitivity of 41 cycles per image. Note that the boundaries do not capture high energy regions of the spectrum.

7.3 Contrast polarity reversal

In this section two phenomena associated with opposite contrast Glass patterns will be considered. The first is the reversal of flow direction in patterns composed of dipoles containing opposite contrast dots: psychophysical data taken from Kovacs and Julesz (1992) are compared to predictions from the adaptive filtering model. The second is the perception of correct flow direction in patterns composed of two superimposed opposite contrast Glass patterns (Prazdny, 1986a). Here a demonstration is given incorporating a simple non-linear transformation of the input image.

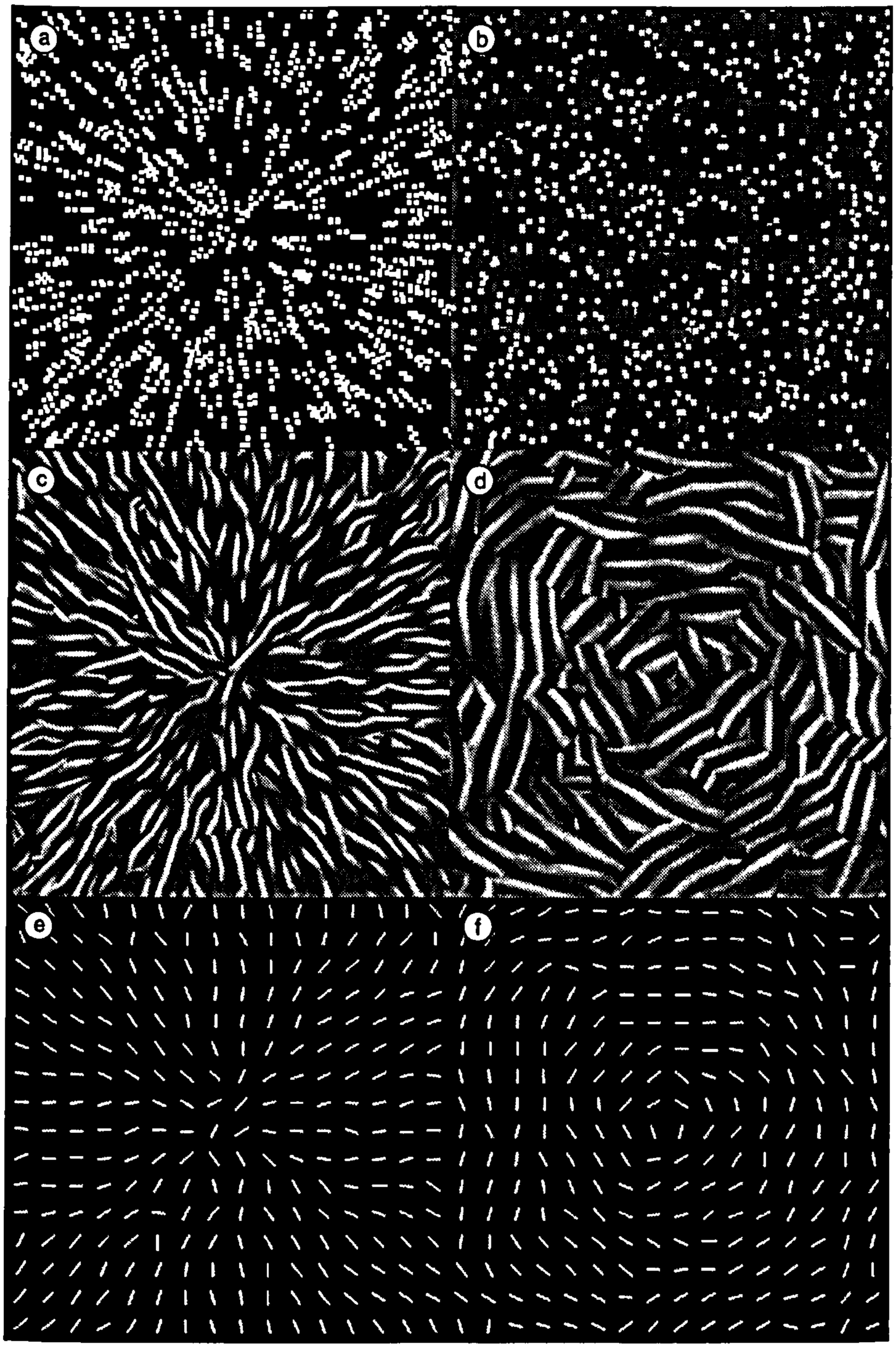


Figure 7.5. The effect of reversing contrast polarity on the model. (a-b) Same and opposite contrast polarity Glass patterns. (c-d) Above patterns filtered at the optimal spatial scale selected by the model. (e-f) The derived flow fields. Note that the model predicts a reversal of local flow pattern in (b).

7.3.1 Pattern reversal: Kovacs and Julesz (1992)

An opposite contrast Glass pattern is produced in the same way as an ordinary pattern except that the first set of randomly distributed features is of opposite contrast polarity to the second, transformed set[2]. An example is shown in Figure 7.4a. The failure to see horizontal structure in this pattern is explicable in terms of the Fourier power spectrum presented in Figure 7.4b. There is very little horizontal energy at any spatial scale, indicated by the dark vertical strip centred on the origin.

Prazdny (1986b) has claimed that correct structure cannot be seen in opposite contrast Glass patterns, but the impression gained, from a circular pattern for example (Figure 7.5a), is usually of a "petal" or "spiral" pattern[3] (Glass and Switkes, 1976). Glass and Switkes (1976) have stated that this percept is inconsistent with a physiological model based on summation of inputs from Kuffler-type receptive fields. This is because simple cells sum the input from either on-centre or off-centre fields but not from both (Hubel and Wiesel, 1967; Tolhurst and Thompson, 1975). Thus opposite contrast dots should not activate simple cells sensitive to the orientation of the dipole and so, according to Glass and Switkes (1976), "the hypothesised physiological mechanism does not appear to explain the observation of spiral-like patterns".

These spirals may, however, simply be considered as a partial rotations of perceived local flow direction. It follows that, under given certain dot densities/separations, flow direction might be *orthogonal* to the original orientation. Exactly such a percept has recently been described and examined by Kovacs and Julesz (1992). They systematically varied the background luminance of horizontal and vertical translational Glass patterns, as a function of the mean luminance of the dots making up each dipole. Given a pair of dots with contrasts C_{pos} and C_{neg} (where $C_{neg} < C_{pos}$), there are three critical ranges of the background luminance, C_{back}:

1. $C_{neg} > C_{back}$ or $C_{pos} < C_{back}$ (A similar contrast polarity pattern).

2. $C_{neg} = C_{back}$ or $C_{pos} = C_{back}$ (A random pattern, since only one feature set is visible).

3. $C_{neg} < C_{back} < C_{pos}$ (A reversed contrast polarity pattern).

---

[2]In the examples shown in the text the order of the opposite contrast elements is also randomised within each dipole.

[3]Anstis (1970) has shown an analogous phenomenon with random dot cinematograms.

Kovacs and Julesz (1992) measured subjects accuracy of reporting when a Glass pattern was a horizontal or a vertical translation. Data from two subjects and two dot luminances from the study (unfilled symbols in Figure 7.6) show that performance approached 100% correct for background contrasts producing features of the same polarity (range 1). As background luminance approached the luminance of one dot set (i.e. range 2), performance fell to the 50% level, as expected. However, as background luminance approached the mean of the feature luminances (i.e. opposite contrast patterns - range 3), performance fell *below* the 50% level, indicating a reversal of perceived orientation (henceforth termed *polarity reversal*). Kovacs and Julesz (1992) go on to show that this phenomenon (unlike analogously constructed random dot cinematograms) cannot be overridden by making matched features the same colour, and suggested that it is form perception which is damaged by the lack of luminance cues offered by these patterns.

What is the explanation for polarity reversal in terms of the flow model described? Figure 7.5 shows two Glass patterns, composed of dot pairs made up of elements of (a) the same and (b) opposite contrast. The scale selection criterion, described earlier, produces the output shown beneath each pattern. Note that the spatial scale of Figure 7.5d is coarser than for Figure 7.5c and the region of integration is correspondingly larger. This scale seems to depend on the density of the reverse polarity pattern: the denser the pattern, the finer the scale selected and the stronger the impression of reversal (Kovacs and Julesz (1992) used very high density (50%) Glass patterns).

Figures 7.5e-f show the flow fields derived from Figures 7.5c-d respectively. The local orientation of 7.5f is clearly locally orthogonal to 7.5e in all regions of the texture. This is not surprising when considering how the presence of a dipole affects the local statistics of a texture. A same-polarity dipole increases the probability that in moving in the direction of the dipole orientation one will encounter an element of the same brightness, and it is this correlation that a Laplacian-of-Gaussian filter highlights. In the case of opposite contrast pairs, the probability of encountering elements of similar brightness in the direction of dipole orientation has been greatly *reduced*. Another way of stating this is, to say that the relative probability that one will encounter elements of similar brightness in all other directions is increased. It is this correlation that produces more energy in coarser scales orthogonal to dipole orientation.

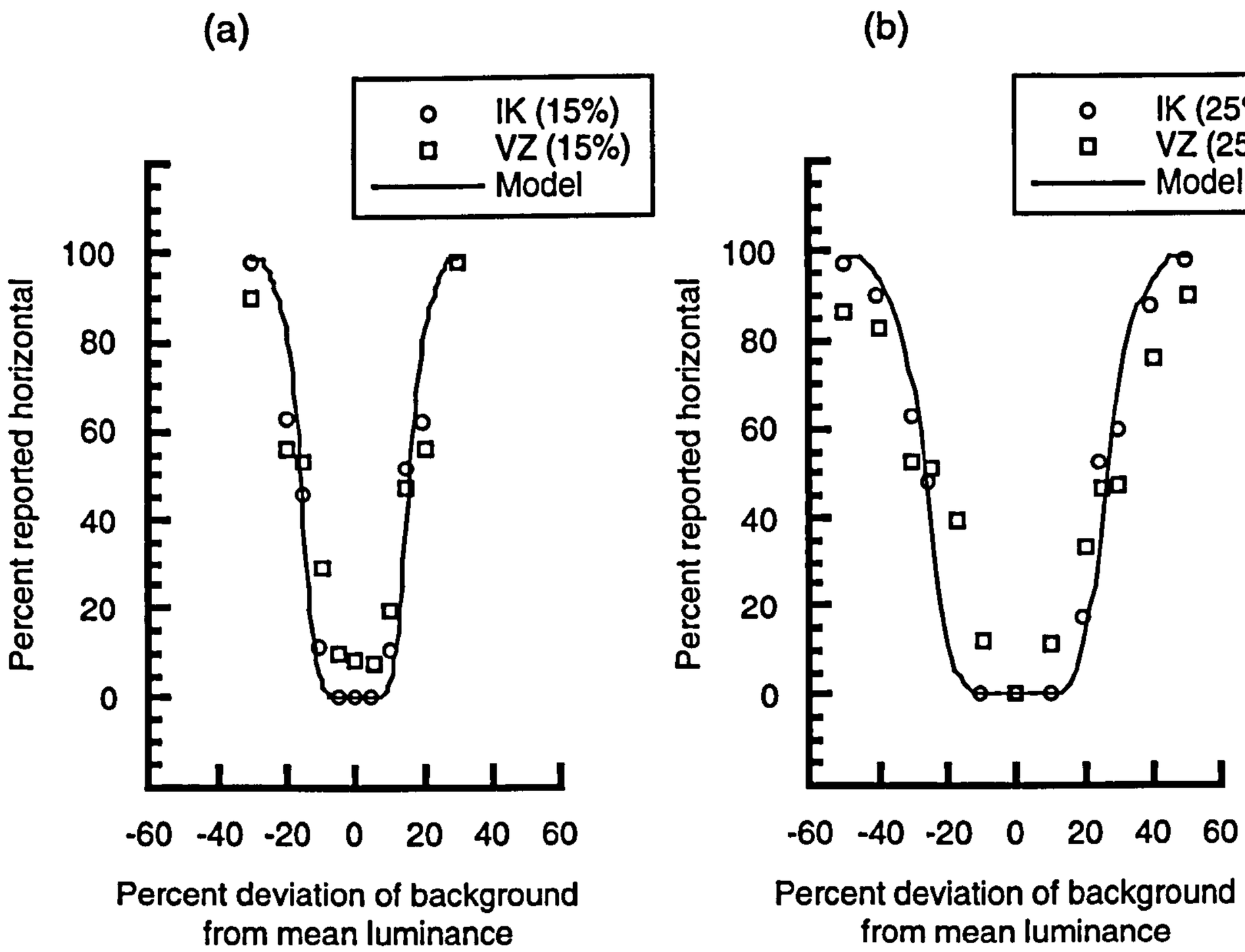


Figure 7.6. Unfilled symbols show the percentage of Glass patterns classified as horizontal as a function of background luminance, for two subjects, at two levels of mean luminance of correlated features (data are taken from Kovacs and Julesz (1992)). The solid lines are predictions from the adaptive filtering model.

Simulation

In order to compare the model's performance on patterns with different background luminances to human data, a Monte Carlo simulation of the Kovacs and Julesz (1992) task was performed. Glass patterns were generated which were similar to the stimuli used in Kovacs and Julesz (1992); each pattern was a 90 X 90 pixel, 256 grey level image. All patterns were composed of a two pixel horizontal translation (i.e. the distance from the centre of an original pixel to the centre of a translated pixel was three pixels). One hundred patterns were generated at 40 levels of background luminance, which ranged from $-2C_{neg}$ to $2C_{pos}$ in steps of $(C_{pos} - C_{neg})/20$. The model, exactly as described in the previous chapter[4], was used to make an estimate of the mean orientation of each pattern, in the range $0 - 180°$. If the reported orientation was between 45° and 135°, the pattern was classified as vertical, otherwise it was classified as horizontal.

Results

The solid lines in Figure 7.6 show the percentage of patterns classified as horizontal as a function of the background luminance, compared to human performance (unfilled symbols). Predictions from a single simulation have been matched to data from the two dot luminance conditions by scaling the physical luminances used in the simulations to the appropriate units used in Kovacs and Julesz (1992) but no fitting has been applied to the data. Note that the model shows good agreement with the human subjects at the three contrast ranges of primary interest (χ^2 measures of the fit are not reported because the standard error of human data were not available). The model reports 100% correct for same polarity patterns, around 50% when only one feature set is visible, and 0% correct (i.e. full polarity reversal) when pure opposite contrast patterns are presented. This is taken as good evidence that reversal of pattern direction is quantitatively explicable in terms of the filtering model described.

7.3.2 Interleaved patterns: Prazdny (1986a)

As noted previously, Prazdny (1986a) has made two criticisms of explanations of grouping in Glass patterns based on spatial filtering. The first relates to how filter size is automatically selected, which has been considered extensively in the two previous chapters. The second is how to deal with a specific type of Glass pattern, composed of two "interleaved"

[4]The model again averaged blob orientations over the whole pattern.

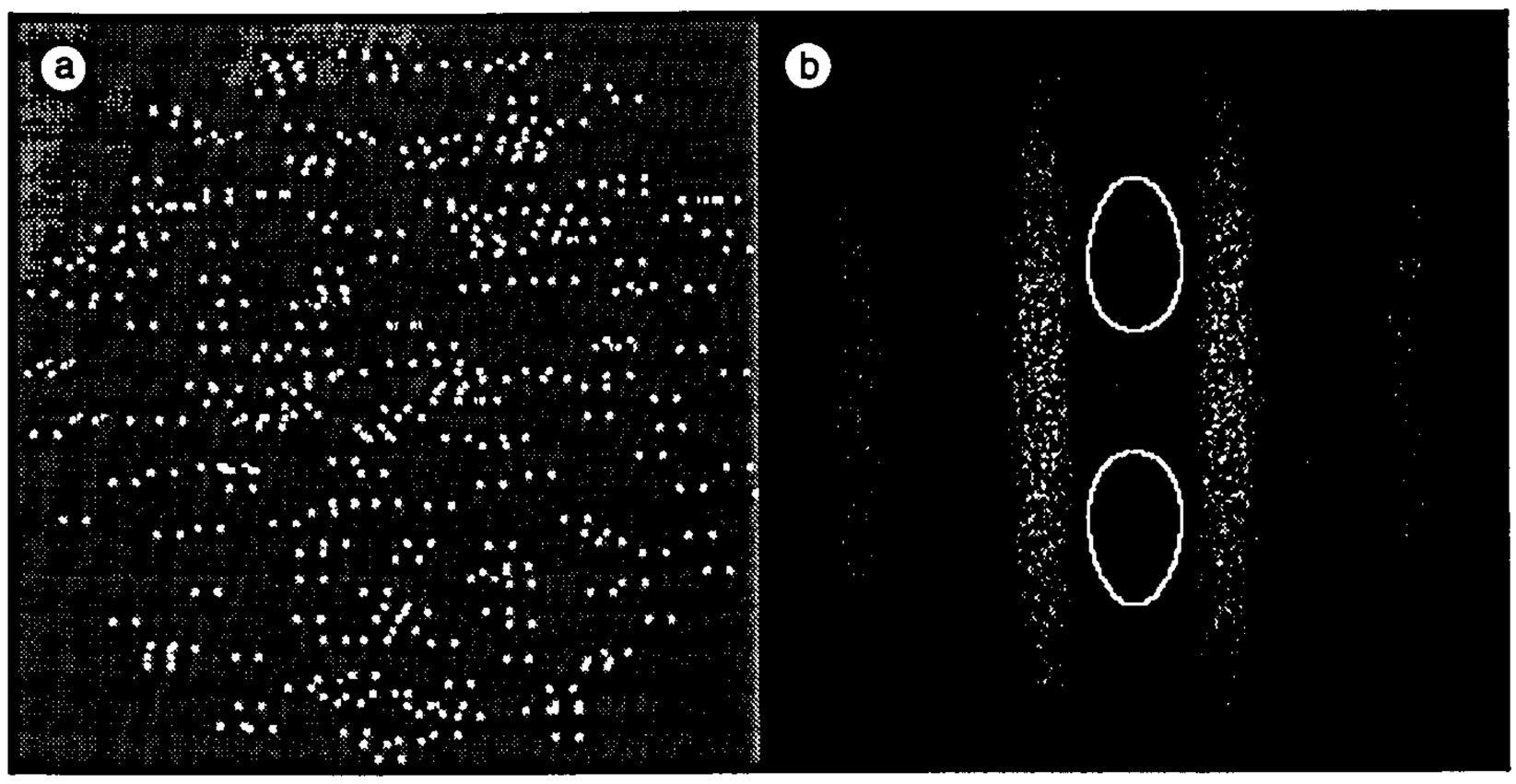


Figure 7.7. (a) Horizontal Glass pattern, composed of alternating opposite contrast dots, with a dot separation of four dots. (b) Fourier power spectrum of (a). Superimposed boundaries show the approximate limit of the sensivity of a DoG with peak sensitivity of 41 cycles. Note that the boundaries capture low energy portions of the spectrum.

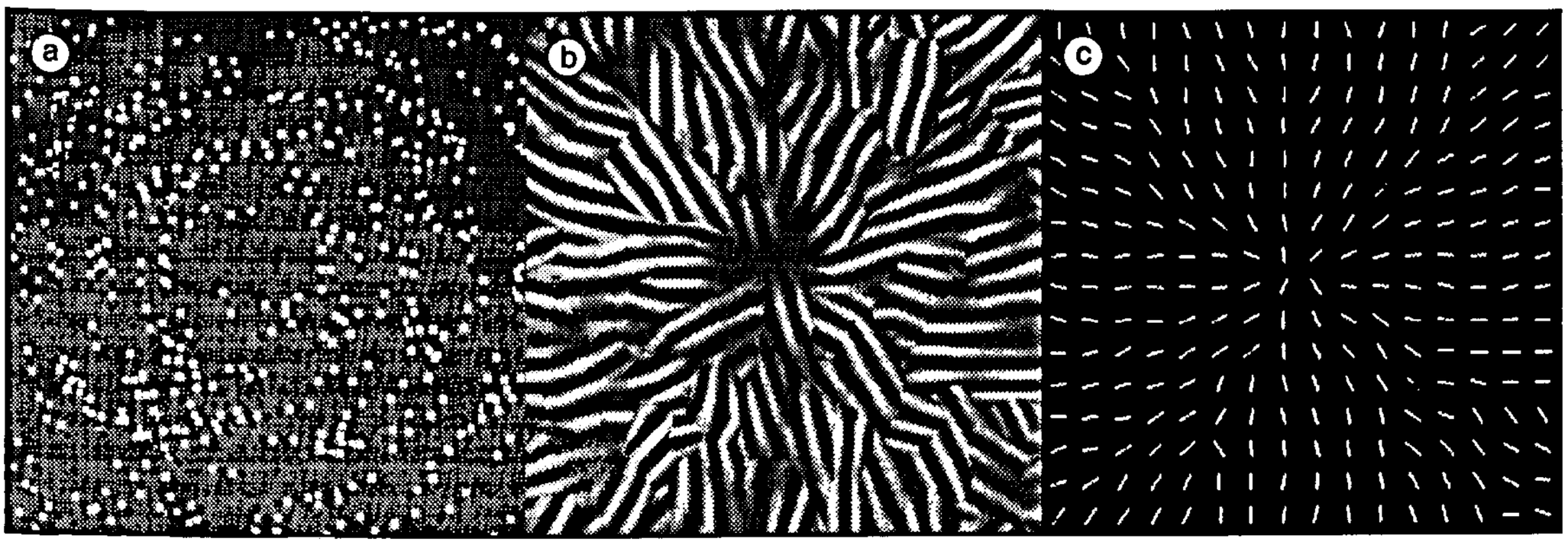


Figure 7.8. (a) An interleaved Glass pattern and (b-c) the result of running it through the oriented filter model. (b) Shows the optimal spatial scale and (c) the resulting flow field.

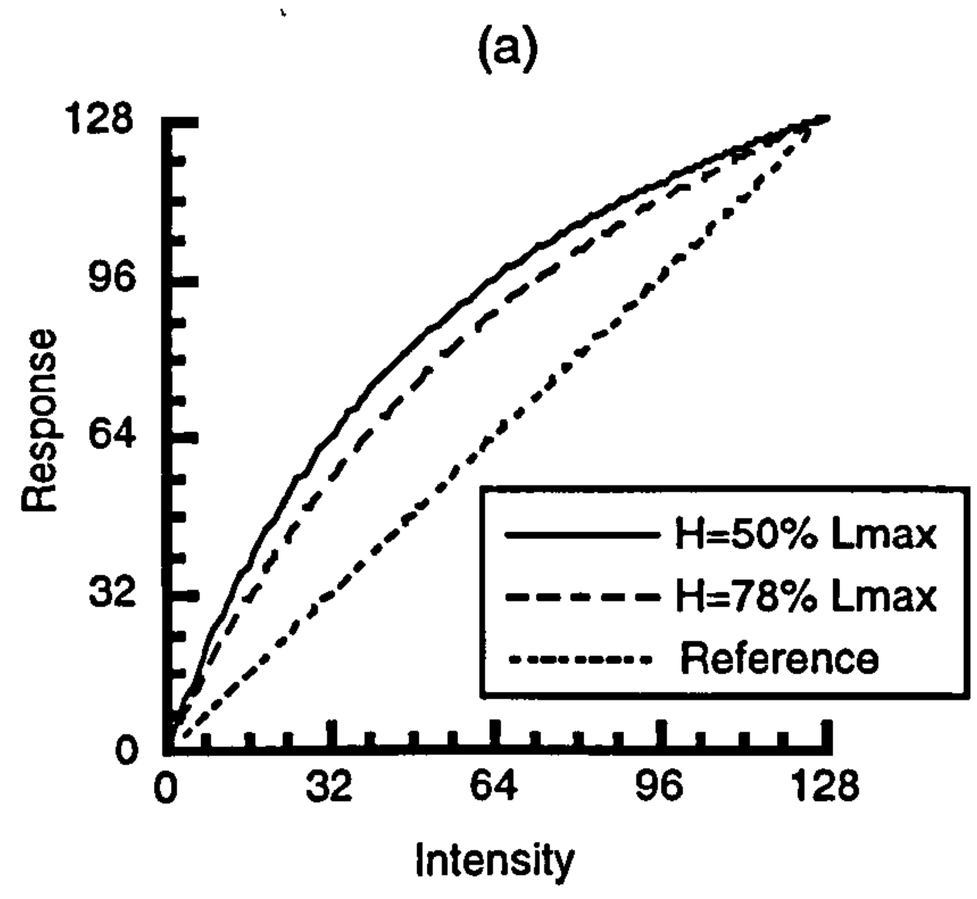


Figure 7.9. The H transform. The x axis shows the image intensity and the y axis the output of the H transform. The transform responses have been normalised to the (original) $0-128$ range.

patterns of opposite contrast. Figure 7.7a shows such a pattern where each texture element is composed of four dots with alternating contrast polarity. Subjectively, at least, textures like these do not produce the pattern reversal shown for the ordinary opposite contrast patterns. Prazdny (1986a), however, states that the only structure visible in band-pass filtered versions of such patterns lies orthogonal to the perceived direction of flow, which makes the explanation of correctly perceived structure, in terms of linear filtering, problematic. The power spectrum of an interleaved pattern, shown in Figure 7.7b, confirms that there is the amount of energy in the horizontal direction across scale, indicated by the faint vertical stripes around the image origin, is very small. Figures 7.8b-c confirm that the adaptive filtering model does indeed produce a reversal of flow direction when presented with such a pattern. This is because these patterns can be thought of as balanced patterns that have been specifically formulated to disrupt oriented filtering mechanisms.

Non-linear transduction of image intensity

One possible explanation for perception of correctly oriented structure in these patterns is that our analysis so far has made a rather sweeping assumption, namely that, at most, a linear transformation of image luminance occurs before filtering. There is considerable evidence from psychophysics (e.g. Legge and Kersten, 1983; Morgan, Mather, Moulden and Watt, 1984) and neurophysiology (e.g. Naka and Rushton, 1966; Shapley and Victor, 1979) that this is not the case. Instead it has been proposed that a compressive non-

linearity modifies intensity before filtering. Indeed the utility of an early non-linearity has already been recognised in a number of models of texture segregation motivated by human psychophysical performance (Graham *et al.*, 1992). Different types of non-linearity have been proposed, including the sigmoid (Gelb and Wilson, 1983), and half-wave rectification (Heeger, 1991), but a particularly computationally simple formulation is the *H-transform* (Naka and Rushton, 1966; Morgan *et al.*, 1984; Watt, 1991b), which has the form:

$$L'(x,y) = L(x,y)\frac{L_{max}}{H + L(x,y)}$$

where $L(x,y)$ is the image intensity at position (x,y), $L'(x,y)$ is the output of the H transform at that point, L_{max} is the maximum image intensity, and H is a constant.

The operation of the H transform is shown in Figure 7.9 which illustrates its effect of compressing high image intensities into a smaller band of responses leaving a larger range for the lower intensities. The value of $H = 0.78L_{max}$ is a typical value derived from a series of psychophysical (edge alignment) tasks (Morgan *et al.*, 1984). Though such a non-linearity is more compressive than that suggested by the light adaptation process in the retina (e.g. Sperling and Sondhi, 1968), there is evidence that such a severe non-linearity is consistent with the function of M-cells in the lateral geniculate nucleus (e.g. Derrington and Lennie, 1984; Sclar *et al.*, 1990).

The result of applying such a non-linearity to the interleaved patterns described above is to shift the luminance of the background towards that of the bright feature set. Thus the bright features will interfere less with the filters' grouping of the dark set. Figure 7.10 shows the result of running the flow algorithm on an interleaved pattern which has been H transformed with values of $H = 0.78L_{max}$ and $H = 0.50L_{max}$ (the maximum value of H not giving polarity reversal). It is clear that, for this pattern at least, values of H around $0.5L_{max}$ produce correctly oriented structure from the model. This implies that a rather severe non-linearity is preceding filtering. However referring back to Figure 7.9, the actual pattern of responses from such a non-linearity is not greatly different to that of a transform with $H = 0.78L_{max}$, a value which has been estimated from other work (Morgan *et al.*, 1984).

Note also that Figure 7.10 demonstrates that perceived flow switches abruptly as the severity of the non-linearity is changed; there is no intermediate flow. According to Prazdny (1986b), bistability is a necessary feature of a model of these phenomena and is one that purely linear filtering models do not have. This demonstration suggests that the

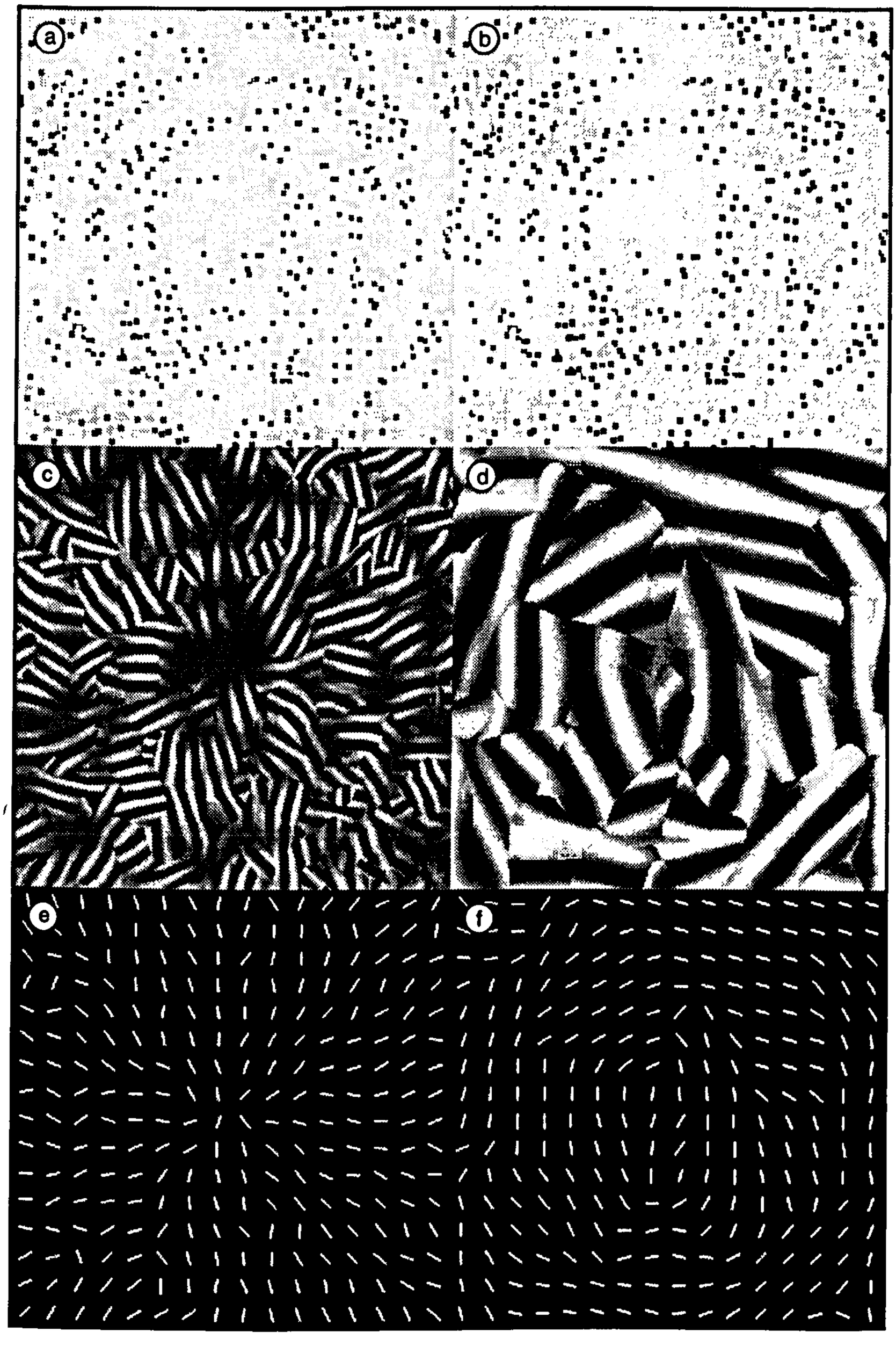


Figure 7.10. The effect of the H transform on the processing of interleaved patterns by the adaptive filtering model. (a) and (b) show the same image H transformed with values of $H = 0.5L_{max}$ and $H = 0.78L_{max}$. The operation of the model on the output of the more severe transformation, shown in the right half of the figure, produces the correct flow field (f).

bistability of these patterns, reported by some observers, is the result of change in this early non-linearity which in turn could be due to something as basic as retinal adaptation, depending on where one locates this non-linearity in the visual system. Figure 7.10 also shows that it is very coarse scale filters which are used to calculate the orientation structure of these patterns. This allows the clear prediction that perception of structure should be significantly impaired, compared to perception of structure in single contrast patterns, in the presence of a spatially low-pass mask, since intermediate scales are of no use for deriving orientation structure.

7.4 Conclusions

The purpose of this chapter was to show that criticisms of spatial filtering models of grouping using contrast phenomena tend to be based on either of two tenuous suppositions: that all visual processing preceding filtering is *entirely linear*, or that the grouping is achieved using *isotropic* filtering. It has been shown that the use of local power to select filter output adaptively, correctly predicts structure in balanced Glass patterns. Through the addition of an early (albeit fairly severe) non-linearity to the adaptive filtering model it has also been shown that the perception of structure in interleaved patterns, which have specifically been constructed as counter-examples of the power of these models at predicting perceived grouping, may be explained in a straightforward manner. The adaptive filtering model correctly predicts that for dense patterns composed of opposite contrast pairs, local flow direction will be rotated by 90°. Psychophysical data from a study by Kovacs and Julesz (1992), quantifying this effect, have also been shown to be consistent with the adaptive filtering model.

A comprehensive explanation of where a non-linearity might be incorporated into the model has not been provided, since it has been suggested that the severity of this function may be affected by retinal adaptation (Hood *et al.*, 1979; Morgan *et al.*, 1984) and by *local* mean luminance (Watt, 1991b). The problems of dealing with these factors go beyond the scope of this thesis. Suffice to say that the addition of a non-linearity seems to provide a natural explanation for Prazdny's (1986a) demonstration and suggests further work on the role of early non-linearities and adaptation in texture perception (Chubb *et al.*, 1989).

8 | Conclusions

In this chapter I will review the findings of this thesis and relate them to the view of texture expounded in Chapter 1, and to some of the specific features of the adaptive filtering model (Chapter 2).

8.1 A statistical representation of texture

The psychophysical experiments using single and mixed orientation distributions (Chapter 3) present the most direct evidence that texture processing involves the estimation of the statistics of local texel attributes. Experiment 3.1 suggested that subjects could only be performing the orientation discrimination task described, by *combining* local orientation estimates over space. This confirms Watt's (1991b) report, and is consistent with the literature examining texture integration over space (e.g. Nothdurft, 1991b; Sagi, 1990). The discrimination of texture containing two orientation distributions, from a field composed of only one set, is determined by the size of the mean orientation offsets of the two sets. This value depends on the *relative* number of elements in each set (Experiment 3.2). This discrimination occurs at threshold mean orientation offsets below the resolution limit of those sets. This indicates that variance information of the unresolved set is probably available to perform this task. When the variance cue is eliminated (by using a relative numerosity judgement in Experiment 3.3) thresholds are higher, and exceed the resolution limit. This suggests that the direction of skew of an orientation distribution is not available. Performance on judging the mean of asymmetrical orientation distributions suggested that the process of combination must incorporate some form of centroid measure to perform this task. However, Experiment 3.6 showed that subjects do not necessarily use the centroid of measures taken over the whole orientation range, but can calculate the centroid over a limited range of orientations, thereby limiting the effects of noise. It was suggested that a possible mechanism for this was an "orientation filter"

operating on the estimated orientation p.d.f. for the texture.

In summary these experiments constrain the statistical representation as follows:

- Combination of orientation measures occurs across space.
- Distinct oriented sets are derived using some form of process to extract *relative* frequency from background noise, such as orientation filtering.
- Resolved orientation distributions are represented independently.
- These sets are represented in terms of their zero-bounded centroid, width (possibly using variance), but not their direction of skew.

The first and last of these constraints are in accord with the adaptive filtering model, but the middle two are not. The adaptive filtering model coded a set of local centroids without making use of orientation filtering because the parameters of such filters were not known. This is evidence that the model should be extended to deal with the representation of multiple local orientations at a point.

Note that these constraints operate at the algorithmic level of a computational theory. Whilst the implementation described in Chapter 2 operates on individually derived texture elements, the evidence presented in Chapter 3 does not constrain the physical realisation of an algorithm sufficiently for other alternatives to be rejected. A model need not actually extract texture elements to perform these tasks. For example a valid alternative might be the use of a model consisting of a set of channels tuned to a range of orientations. However, note that these constraints apply equally in the case of such a model. Average orientation from a channel coded model cannot be derived using the most active orientation channel - a degree of integration across channels must occur. Orientation variance might then be coded as the spread of activity over channels. Similarly, subjects' insensitivity to the d.c. term of orientation distributions, demonstrated in Experiment 3.6, suggests that some form of inter-channel inhibition would also have to be incorporated. Given that the tasks being investigated do not operate at the very lowest levels of visual processing, devising tasks which completely separated the predictions of models, at an implementational level, is probably not feasible. What Chapter 3 describes is a set of texture statistics that any adequate model of texture perception must be able to produce.

8.2 Measurement of textural properties

Chapters 4-5 presented psychophysical evidence relating to the way texture elements are extracted from images. The main finding, with reference to the computational modelling of this process, is that human subjects out-perform models based on isotropic filtering on certain texture tasks (Experiment 4.2). This suggests that in general the accuracy of orientation measures derived from schemes using e.g. Laplacian-of-Gaussian filtering (Voorhees and Poggio, 1987; Witkin, 1981) may not achieve human levels of performance at tasks using orientation information derived from those measures (e.g. segmentation). Instead it was suggested that subjects must have access to some form of oriented filtering. Models using the output of individual DoGs outperform human subjects on the tasks described, but beg the question of how one selects the correct orientation. The adaptive filtering model, by using point-wise selection of the most locally active filter output, gets around this problem but results in a loss of orientation sensitivity, when compared to DoGs. This loss of sensitivity is consistent with human subjects' discrimination of texture orientation.

The linearity of the mechanisms for deriving texture elements was considered in Chapter 7. It was shown that the use of local power to select filter output adaptively correctly predicts structure in Glass patterns composed of high-pass elements. The adaptive filtering model also correctly predicts that for dense patterns, composed of short opposite contrast pairs, local flow direction is rotated by 90°. The model produces quantitative predictions of performance consistent with psychophysical data from a study examining this phenomenon (Kovacs and Julesz, 1992). By adding an early non-linearity to the model, it was also shown that structure can be derived from patterns composed of interleaved opposite contrast patterns (textures specifically constructed to disrupt the output of spatial filtering mechanisms). The purpose of these demonstrations was not to definitively establish the presence of, or to locate, any early non-linearities the visual processing of texture. Rather it was intended to show that studies which attempt to disprove the presence of spatial filtering in early visual grouping must make some highly restrictive assumptions, such as pure linearity.

8.3 Selection of spatial scale

One of the assumptions outlined in Chapter 1 was that certain scales are better than others for processing texture, and that some form of selection must occur between scales.

The results from the modelling of mean orientation discrimination tasks, described in Chapter 5, certainly suggests that this is the case. The scale of DoG and $\nabla^2 G$ filters used to derive orientation estimates from Glass patterns critically effects how accurate those orientation estimates are. The task of deriving a mean orientation estimate naturally constrains scale selection: texture elements should have minimum local orientation variance. All the modelling used for this thesis has employed only that criterion for selecting spatial scale.

The implications from this work is that minimum local orientation variance is a useful general strategy for selecting spatial scale. From Chapter 2 it is suggested that this has a place in systems for deriving shape-from-texture, discontinuities from texture, texture flow, and texture contour analysis.

8.4 Integration across space

This thesis has concentrated largely on the extraction of information from spatially unstructured texture patches. Chapter 6 did consider spatial aspects of the model by considering how globally organised patterns are coded. The filter model combined with a measure of mean local orientation variance, or a measure of the degree of local orientation matching to a template, was used to simulate three psychophysical tasks.

Simulation of data from Maloney *et al.* (1987) (Experiment 6.1) demonstrated that while a measure of local orientation variance could predict the discriminability of Glass patterns in the presence of extra noise dots there had to be sufficient texels present for the measure to be of any use. At low texel densities it was suggested that subjects make use of more detailed knowledge about the underlying transform.

The Hel Or and Zucker (1989) task is closely related to a variety of texture segmentation tasks. The success of the orientation variance model suggests that this task does not require the explicit coding of the *position* of texture boundaries, and that instead local orientation statistics suffice. This is illustrative of the problem of presenting subjects with segmentation tasks - the information they can use is often insufficiently constrained.

In summary the assumptions made about the spatial integration used by the model have been upheld, in the limited tests presented in Chapter 6. More psychophysical work would be required to fully evaluate this aspect of the model. However mean local orientation variance seems to be a good indicator of the presence of structure in texture.

8.5 Glass patterns

The experiments reported in this thesis have made extensive use of Glass patterns as examples of texture which only contain structure at a narrow band of scales. It is proposed that they are an important class of stimuli for examining scale selection because they contain information that actively disrupts estimates of orientation at non-optimal scales. Spatially narrow-band textures, such as filtered noise, are not appropriate because these patterns merely contain less energy at other spatial scales.

The effects of element length (Caelli, 1981; Jenkins, 1983; Wagemans *et al.*, 1993), and the lack of effect of element density (Jenkins (1983)) were confirmed. Local orientation variance was found to have no effect on the judgement of mean orientation until around a s.d. of around 8°, suggesting that there is intrinsic noise on the estimation of mean orientation in this task. DoG and $\nabla^2 G$ filtering models produced good accounts of human data in the density and variance conditions, but only the DoG was adequate to explain the effect of dipole length. The Stevens model (Stevens, 1978) failed to achieve human levels of performance in any of the conditions, even when neighbourhood size was set to exclude the maximum possible number of false matches. It is concluded that this model is unable to make use of the coarse scale interactions that occur when co-linear dipoles are permitted (unlike Stevens, 1978).

Chapter 6 presented three tasks involving the detection of structure in Glass patterns, in order to investigate the representation of structure from patterns which contain systematic orientation changes. A model which measures local orientation standard deviation suffices to explain performance in all but the lowest density conditions of Maloney, Mitchison and Barlow's (1987) experiment. A model calculating local deviations from expected orientation structure, consistently exceeds subjects' performance. Similarly both models account for the threshold amount of orientation jitter that subjects can tolerate in discriminating cued patterns from noise. The final experiment replicated Hel Or and Zucker (1989) and investigated the detection of anomalous patches embedded within oriented Glass patterns. Surprisingly, a model which located regions of the image that were low orientation variance regions failed to produce sufficiently accurate performance to explain human subjects' behaviour. However the use of a simple measure of the mean local orientation variance *did* produce behaviour which was reasonably close to subjects. This task most closely resembles a texture segmentation task and suggests that the model could be appropriate to such tasks. It also suggests that psychophysical tasks investigating the

detection of texture *boundaries* must avoid using tasks that can be achieved using simple global statistics.

To conclude this section, the psychophysical data collected using Glass patterns have been shown to be largely explicable by models based on spatial filtering. The balance of evidence from Chapter 5 suggests that for tasks which place statistical constraints on subjects' performance, a model based on oriented filters, operating at a scale which minimises local orientation variance, suffices to explain performance in all conditions. The use of local orientation variance as a measure of structure for other tasks, was further demonstrated in Chapter 6.

8.6 Further work

This thesis has demonstrated the utility of maximising anisotropy for scale selection in the context of the texture tasks described. Would this criterion be useful for other tasks, such as object recognition? Additionally, what *other* criteria might be used? From the work of Emmott (1994) there seem to be regularities in the output of filters responding to pages of text, that suggest that scale selection for reading could rely on equally simple, statistical strategies.

The application of the adaptive filtering model to contour integration is also worthy of further investigation. As well as the work of Field *et al.* (1993), recent research by Polat and Sagi (1994) is of interest. The contrast threshold for detection of Gabor patches is improved if patches are flanked by co-linear Gabor patches, but not by patches that could not have derived from a common contour. It is possible that the adaptive filtering model could be used to explain this phenomenon.

Adaptive filtering might be used to detect other forms of regularity in dot patterns. It's resistance to noise suggest that it may be appropriate for detecting very long range groupings in patterns, such as might be caused by symmetry, for example.

References

Ajuha, N. and Schachter, B. (1982). *Pattern models.* New York, NY: John Wiley and Sons.

Aloimonos, J. and Swain, M. (1984). Shape from texture. *Proceedings of the 9th International Conference on Artificial Intelligence*, pages 926–931.

Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics*, *58*, 345–360.

Anstis, S. (1970). Phi movement as a subtraction process. *Vision Research*, *10*, 1411–1430.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique" effect in man and animals. *Psychological Bulletin*, *78*, 266–178.

Badcock, D. and Westheimer, G. (1985). Spatial location and hyperacuity: The centre/surround localization contribution function has two substrates. *Vision Research*, *25*, 1259–1267.

Bajczy, R. and Lieberman, L. (1976). Texture gradient as a depth cue. *Computer Graphics and Image Processing*, *5*, 52–67.

Barlow, H. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London*, *B 290*, 71–82.

Beck, J. and Ambler, B. (1972). Discriminability of differences in line slope and in line arrangements as a function of mask delay. *Perception and Psychophysics*, *12*, 33–38.

Beck, J. and Ambler, B. (1973). The effect of concentrated and distributed attention on peripheral acuity. *Perception and Psychophysics*, *14*, 225–230.

Beck, J., Prazdny, K., and Rosenfeld, A. (1983). A theory of textural segmentation. In Beck, J., Hope, B., and Rosenfeld, A. (Eds.), *Human and machine vision*, pages 1–38. London: Academic Press.

Beck, J., Sutter, A., and Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segmentation. *Computer Vision, Graphics and Image processing*, *37*, 299–325.

Beck, J. (1966). Effect of orientation and of shape similarity on shape perception. *Perception and Psychophysics*, *1*, 300–302.

Beck, J. (1967). Perceptual grouping produced by line figures. *Perception and Psychophysics*, *2*, 491–495.

Beck, J. (1972). Similarity grouping and peripheral discriminability under uncertainty. *American Journal of Psychology*, *85*, 1–19.

Beck, J. (1982). Textural segmentation. In Beck, J. (Ed.), *Organisation and representation in perception*, chapter 15, pages 285–317. Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Bergen, J. and Adelson, E. (1988). Early vision and texture perception. *Nature*, *333*, 363–364.

Bergen, J. and Landy, M. (1991). Computational modeling of visual texture segregation. In Landy, M. and Movshon, J. (Eds.), *Computational models of visual processing*, pages 253–271. Cambridge, MA: MIT Press.

Bergen, J. (1991). Theories of visual texture perception. In Regan, D. (Ed.), *Vision and visual dysfunction. Volume 10: Spatial vision*, pages 114–134. London: Macmillan.

Bishop, P., Coombs, J., and Henry, G. (1973). 231. *Journal of Physiology, London*, pages 31–60.

Blake, A. and Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, *45*, 323–380.

Blake, A., Bülthoff, H., and Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, *33*, 1723–1737.

Blostein, D. and Ahuja, N. (1993). Shape from texture: Integrating texture-element extraction and surface estimation. *Vision Research*, *33*, 1723–1737.

Bovik, A., Clark, M., and Geisker, W. (1990). Multi-channel texture analysis using localised spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*, 55–73.

Burr, D., Morrone, M., and Maffei, L. (1981). Intracortical inhibition prevents simple cells from responding to textured patterns. *Experimental Brain Research*, *43*, 455–458.

Caelli, T. and Julesz, B. (1979). Psychophysical evidence for global feature processing in visual texture discrimination. *Journal of the Optical Society of America*, *A69*, 675–678.

Caelli, T., Preston, G., and Howell, E. (1978). Implications of spatial summation models for processes of contour perception. *Vision Research*, *18*, 723–734.

Caelli, T. (1981). Some psychophysical determinants of discrete Moiré patterns. *Biological Cybernetics*, *39*, 97–103.

Caelli, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, *1*, 19–30.

Campbell, F. and Robson, J. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, *197*, 551–566.

Carlson, C., Anderson, C., and Moeller, J. (1980). Visual illusions without low spatial frequencies. *Investigative Opthamology and Visual Science: Supplement*, *19*, 165.

Chubb, C. and Landy, M. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In Landy, M. and Movshon, J. (Eds.), *Computational models of visual processing*, chapter 19, pages 291–301. Cambridge, MA: MIT Press.

Chubb, C., Sperling, G., and Solomon, J. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Science, USA*, *86*, 9631–9635.

Coggins, J. and Jain, A. (1985). A spatial-filtering approach to texture analysis. *Pattern Recognition Letters*, *3*, 195–203.

Cumming, B., Johnston, E., and Parker, A. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*, 827–838.

Cutting, J. and Millard, R. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198–216.

Daugman, J. (1980). Two dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*, 847–856.

Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimised by two-dimensional cortical filters. *Journal of the Optical Society of America, 2*, 1160–1169.

Davis, L., Janos, L., and Dunn, S. (1983). Efficient recovery of shape from texture. *Pattern Analysis and Machine Intelligence, 5.*

Derrington, A. and Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 219–240.

Dodwell, P. and Caelli, T. (1985). Recognition of vector patterns under transformation: Local and global determinants. *Quarterly Journal of Experimental Psychology, 37A*, 1–23.

Emmott, S. (1994). *The visual processing of text.* PhD thesis, Stirling: University of Stirling.

Enns, J. (1986). Seeing textons in context. *Perception and Psychophysics, 39*, 143–147.

Field, D., Hayes, A., and Hess, R. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research, 33*, 173–193.

Fogel, I. and Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics, 61*, 103–113.

Fu, K. (1974). *Syntactic methods in pattern recognition.* New York, NY: Academic Press.

Garcia-Perez, M. (1991). Visual phenomena without low spatial frequencies: A closer look. *Vision Research, 31*, 1647–1653.

Gårding, J. (1993). Direct estimation of shape from texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*, 1202–1208.

Gelb, D. and Wilson, H. (1983). Shifts in perceived size as a function of contrast and temporal modulation. *Vision Research, 23*, 71–82.

Gibson, J. (1950a). *The perceptino of the visual world.* Boston, MA: Houghton Mifflin.

Gibson, J. (1950b). *The perception of the visual world.* Boston, MA: Houghton Mifflin.

Gibson, J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Gilbert, C. and Wiesel, T. (1989). Columnar specificity of intrinsic horizontal and corticocortcal connections in cat visual cortex. *Journal of Neuroscience, 9,* 2432–2442.

Glass, L. and Perez, R. (1973). Perception of random dot interference patterns. *Nature, 246,* 360–362.

Glass, L. and Switkes, E. (1976). Pattern recognition in humans: Correlations which cannot be perceived. *Perception, 5,* 67–72.

Glass, L. (1969). Moiré effects from random dots. *Nature, 243,* 578–580.

Graham, N., Beck, J., and Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research, 32,* 719–743.

Grinvald, A., Tso, D., Frostig, R., Lieke, E., Arieli, A., and Hildesheim, R. (1989). Optical imaging of neuronal activity in visual cortex. In Beck, J., Hope, B., and Rosenfeld, A. (Eds.), *Neural mechanisms of visual perception,* pages 117–135. Woodlands, Texas: Portfolio.

Gurnsey, R. nd Browse, R. (1987). Micropatterns properties and resentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 41,* 239–252.

Hallett, P. (1992). Segregation of mesh-derived textures evaluated by resistance to added disorder. *Vision Research, 32,* 1899–1911.

Harvey, L. and Gervais, M. (1978). Visual texture perception and Fourier analysis. *Perception and Psychophysics, 24,* 534–542.

Harvey, L. and Gervais, M. (1981). Internal representation of visual texture as a basis for the judgment of similarity. *Journal of Experimental Psychology: Human Perception and performance, 7,* 741–753.

Heeger, D. (1991). Nonlinear model of neural responses in cat visual cortex. In Landy, M. and Movshon, J. (Eds.), *Computational models of visual processing,* pages 119–151. Cambridge, MA: MIT Press.

Hel Or, Y. and Zucker, S. (1989). Texture fields and texture flows: Sensitivity to differences. *Spatial Vision, 4*, 131–139.

Hood, D., Finkelstein, M., and Buckingham, E. (1979). Psychophysical tests of the response function. *Vision Research, 19*, 401–406.

Horn, B. (1986). *Robot vision.* Cambridge, MA: MIT Press.

Hubel, D. and Wiesel, T. (1967). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, pages 106–154.

Ikeuchi, K. (1984). Shape from regular patterns. *Artificial Intelligence, 22*, 49–75.

Janez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research, 24*, 271–274.

Jenkins, B. (1983). Spatial limits to the detection of transpositional symmetry in dynamic dot textures. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 258–269.

Jenkins, B. (1985). Orientational anisotropies in the human visual system. *Perception and Psychophysics, 37*, 125–134.

Julesz, B., Frisch, H., Gilbert, E., and Victor, J. (1973). Inability of human to discriminate between visual textures that agree in second-order statistics - revisited. *Perception, 2*, 391–405.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory, 8*, 84–92.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*, 91–97.

Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics, 84*, 245–251.

Julesz, B. (1991). Early vision and focal attention. *Reviews of modern physics, 63*, 735–772.

Kanatani, K. (1984). Detection of surface orientation and motion from texture by a stereological technique. *Artificial Intelligence, 23*, 213–237.

Kass, M. and Witkin, A. (1985). Analyzing oriented patterns. In *Proceedings of the International Joint Conference on Artificial Intelligence.*

Keeble, M. and Morgan, M. (1993). The perception of oriented textures: An orientational-filter model. *Perception, 22 Supplement,* 70.

Koenderink, J. (1990). *Solid shape.* Cambridge, MA: MIT Press.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York, NY: Harcourt Brace.

Kovacs, I. and Julesz, B. (1992). Depth, motion, and static-flow perception at metaisoluminant color contrast. *Proceedings of the National Academy of Science, 89.*

Landy, M. and Bergen, J. (1991). Texture segregation and orientation gradient. *Vision Research, 31,* 679–691.

Legge, G. and Kersten, D. (1983). Light and dark bars: Contrast discrimination. *Vision Research, 23,* 473–483.

Malik, J. and Perona, P. (1990). Preattentive texture discrimination with early visual mechanisms. *Journal of the Optical Society of America, A7,* 923–932.

Maloney, R., Mitchison, G., and Barlow, H. (1987). Limit to the detection of Glass patterns in the presence of noise. *Journal of the Optical Society of America, A4,* 2336–2341.

Marr, D. (1976). Early processing of visual information. *Proceedings of the Royal Society of London, B 275,* 483–534.

Marr, D. (1982). *Vision.* San Francisco, CA: Freeman.

Moraglia, G. (1989). Display organisation and the detection of horizontal line segments. *Perception and Psychophysics, 45,* 265–272.

Morgan, M. and Fahle, M. (1992). Effects of pattern density upon displacement limits for motion detection in random binary luminance patterns. *Proceedings of the Royal Society of London, 248,* 189–198.

Morgan, M., Mather, G., Moulden, B., and Watt, R. (1984). Intensity-response nonlinearities and the theory of edge localization. *Vision Research, 24,* 713–719.

Morrone, M. and Burr, D. (1986). Evidence for the existence and development of visual inhibition in humans. *Nature*, *321*, 235–237.

Morrone, M., Burr, D., and Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. *Proceedings of the Royal Society of London, B 216*, 335–354.

Mostafavi, H. and Sakrison, D. (1976). Structure and properties of a single channel in the human visual system. *Vision Research*, pages 957–968.

Naka, K. and Rushton, W. (1966). S-potentials from luminosity units in the retina of fish (cyprinidae). *Journal of Physiology, London*, *185*, 587–599.

Nothdurft, H. (1985a). Orientation sensitivity and texture segmentation in patterns with different line orientations. *Vision Research*, *25*, 551–560.

Nothdurft, H. (1985b). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, *25*, 1957–1968.

Nothdurft, H. (1990). Texture segregation by associated differences in global and local luminance distribution. *Proceedings of the Royal Society of London*, *B239*, 295–320.

Nothdurft, H. (1991a). Different effects from spatial frequency masking in texture segmentation and texture detection tasks. *Vision Research*, *31*, 299–320.

Nothdurft, H. (1991b). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*, 1073–1078.

Nothdurft, H. (1992). Feature analysis and the role of similarity in preattentive vision. *Perception and Psychophysics*, *52*, 355–375.

Palmer, S. (1992). Modern theories of Gestalt perception. In Humphreys, G. (Ed.), *Understanding vision*, pages 39–70. Oxford: Basil Blackwell Ltd.

Paradiso, M. (1988). T theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, *58*, 35–49.

Parent, P. and Zucker, S. (1989). Trace inference, curvature consistency and curve detection. *IEEE Transactions on Pattern analysis and Machine Intelligence*, *11*, 823–839.

Parker, A. and Hawken, M. (1988). Two-dimensional spatial structure of receptive fields in monkey striate cortex. *Journal of the Optical Society of America, A5*, 598–605.

Phillips, G. and Wilson, H. (1983). Orientation bandwidths of spatial mechanisms measures by masking. *Journal of the Optical Society of America, A1*, 226–232.

Polat, U. and Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34*, 73–78.

Prazdny, K. (1986a). Psychophysical and computational studies of random-dot Moiré patterns. *Spatial Vision, 1*, 231–242.

Prazdny, K. (1986b). Some new phenomena in the perception of Glass patterns. *Biological Cybernetics, 53*, 153–158.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical recipes in C: The art of scientific computing.* Cambridge: Cambridge University Press.

Reed, T. and Wechsler, H. (1990). Segmentation of textured images and Gestalt organisation using spatial/spatial-frequency representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 1–12.

Rentschler, I., Hubner, M., and Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Research, 28*, 279–291.

Resnikoff, H. (1989). *The illusion of reality.* New York, NY: Springer-Verlag.

Richards, W. and Pollit, A. (1988). Texture matching. *Kybernetik, 28*, 279–291.

Robson, J. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America, 56*, 1141–1142.

Sagi, D. and Julesz, B. (1987). Short range limitations on detection of feature differences. *Spatial Vision, 2*, 39–49.

Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: Effect of signal density and spatial-frequency. *Vision Research, 30*, 1377–1388.

Sclar, G., Maunsell, J., and Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30*, 1–10.

Sedgwick, H. (1983). Space perception. In Boff, K., Kaufman, L., and Thomas, J. (Eds.), *Handbook of perception and human performance:. Volume I: Sensory processes and perception*, pages 21–1 to 21–57. New York, NY: John Wiley and Sons.

Shapley, R. and Victor, J. (1979). The contrast gain control of the cat retina. *Vision Research, 19*, 431–434.

Sperling, G. and Sondhi, M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America, 58*, 1133–1145.

Stevens, K. and Brookes, A. (1987). Detecting structure by symbolic constructions on tokens. *Computer Vision, Graphics and Image Processing, 37*, 238–260.

Stevens, K. (1978). Computation of locally parallel structure. *Biological Cybernetics, 6*, 19–28.

Sutter, A., Beck, J., and Graham, N. (1988). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. Technical Report 82-8, Center for Automation research, Computer Vision Laboratory, University of Maryland, College Park.

Swets, J. (1964). *Signal detection and recognition by human observers.* New York, NY: John Wiley and Sons.

Todd, J. and Akerstrom, R. (1987). Perception of three-dimensional from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and performance, 13*, 242–255.

Tolhurst, D. and Thompson, P. (1975). Orientation illusions and aftereffects: Inhibition between channels. *Vision Research, 15*, 967–972.

Tomita, F., Shirai, Y., and Tsuji, S. (1987). Description of textures by a structural analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-4*, 679–698.

Treisman, A. and Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*, 15–48.

Tsumoto, T., Eckart, W., and Creutzfeldt, O. (1979). Modification of orientation sensitivity of cat visual cortex neurons by removal of GABA-mediated inhibition. *Experimental Brain Research, 34*, 351–363.

Uttal, W. (1975). *An autocorrelation theory of form detection.* Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.

Vassilev, A., Simeonova, B., and Zlatkova, M. (1981). Orientation acuity at detection threshold. In *Proceedings of the Fourth European Conference on Visual Perception*, page 17.

Vilnrotter, H., Nevatia, R., and Price, K. (1986). Structure analysis of natural textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8*, 679–698.

Voorhees, H. and Poggio, T. (1987). Detecting textons and texture boundaries in natural images. In *Proceedings of the First International Conference on Computer Vision*, pages 250–258.

Wagemans, J., Van Gool, L., Swinnen, V., and Van Horebeek, J. (1993). Higher-order structure in regularity detection. *Vision Research, 33*, 1067–1088.

Watt, R. and Andrews, D. (1981). APE: Adaptive probit estimation of psychometric functions. *Current Psychological Review, 1*, 205–214.

Watt, R. and Morgan, M. (1983a). Mechanisms responsible for the assessment of visual location: Theory and evidence. *Vision Research, 23*, 97–109.

Watt, R. and Morgan, M. (1983b). The recognition and representation of edge blur: Evidence for spatial primitives in human vision. *Vision Research, 23*, 1457–1477.

Watt, R. and Morgan, M. (1984). Spatial filters and the localization of luminance changes in human vision. *Vision Research, 24*, 1387–1397.

Watt, R. and Morgan, M. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*, 1661–1674.

Watt, R. (1988). *Visual Processing: Computational, psychophysical and cognitive research.* London: Lawrence Erlbaum Associates Ltd.

Watt, R. (1991a). Seeing texture. *Current Biology, 1*, 137–139.

Watt, R. (1991b). *Understanding vision.* London: Academic Press.

Wen, W. and Fryer, R. (1991). Texture boundary detection - a structural approach. In Mowforth, P. (Ed.), *British Machine Vision Conference 1991*, pages 104–110.

Westheimer, G. (1981). Visual hyperacuity. *Progress in Sensory Physiology*, *1*, 1–20.

Whitaker, D. and Walker, H. (1988). Centroid evaluation in the vernier alignment of dot clusters. *Vision Research*, *28*, 777–784.

Wilson, H. and Gelb, D. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America*, *A1*, 124–131.

Wilson, H. and Regan, D. (1984). Spatial-frequency adaptation and grating discrimination: Predictions of a line-element model. *Journal of the Optical Society of America*, *A1*, 1091–1096.

Wilson, H., McFarlane, D., and Phillips, G. (1983). Spatial tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*, 873–882.

Witkin, A. and Tennenbaum, J. (1983). On the role of structure in vision. In Beck, J., Hope, B., and Rosenfeld, A. (Eds.), *Human and machine vision*, pages 481–543. London: Academic Press.

Witkin, A. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, *17*, 17–47.

Zucker, S. and Davis, S. (1988). Points and enpoints: A size/spacing constraint for dot grouping. *Perception*, *17*, 229–247.

Zucker, S., Stevens, K., and Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception and Psychophysics*, *34*, 229–247.

Zucker, S. (1982). Early orientation selection and grouping: Evidence for type I and type II processes. Technical Report 82-8, Department of Electrical Engineering, McGill University.